

Contents

| | |
|--|-----------|
| 1 Biology—Physiology | 4 |
| 1.1 Types of Tissues | 4 |
| 1.2 Action Potentials | 4 |
| 1.3 The Neuromuscular Junction | 6 |
| 1.4 The Sarcomere | 7 |
| 1.5 Nervous System Components | 7 |
| 1.6 Control of Body Activity and the Neurovisceral Control | 10 |
| 1.7 Receptors and Sensory Input | 11 |
| 1.8 Cardiovascular Anatomy | 11 |
| 1.9 Lymphatic System | 13 |
| 1.10 Blood Clotting | 14 |
| 1.11 The Lungs | 15 |
| 1.12 The Gastrointestinal Tract | 15 |
| 1.13 The Kidney | 17 |
| 1.14 Homeostatic Mechanisms | 19 |
| 1.15 Male Reproduction | 19 |
| 1.16 Female Reproduction | 21 |
| 1.17 Development | 23 |
| 1.18 Developmental Mechanisms | 26 |
| 1.19 Human Embryo Development | 26 |
| 1.20 Endocrinology | 27 |
| 1.21 The Pituitary Gland | 28 |
| 1.22 Immunology | 29 |
| 1.23 Antibodies | 29 |
| 1.24 T Cells | 31 |
| 1.25 Immunology—Generalized Review | 31 |
| 2 General Chemistry | 32 |
| 2.1 Stoichiometry | 32 |
| 2.2 Atomic Theory | 33 |
| 2.3 Periodic Trends | 35 |
| 2.4 Important Periodic Families | 36 |
| 2.5 Electrochemistry | 37 |
| 2.6 Gases and Gas Laws | 38 |
| 2.7 Phases and Phase Changes | 40 |
| 2.8 Colligative Properties | 40 |
| 2.9 Solubility | 41 |
| 2.10 Acids and Bases | 42 |
| 2.11 Titration Curves | 43 |
| 2.12 Equilibrium | 44 |
| 2.13 Thermochemistry | 45 |
| 2.14 Chemical Kinetics | 46 |
| 3 Organic Chemistry of Biological Systems | 46 |
| 3.1 Molecular Structure | 46 |
| 3.2 Isomers and Stereochemistry | 48 |
| 3.3 Nucleophilic Substitution | 50 |
| 3.4 Structure Elucidation | 52 |
| 3.5 Lab Techniques | 55 |
| 3.6 Lipids | 59 |
| 3.7 Lipid Signal Molecules | 60 |
| 3.8 General Bioorganic Reactions | 62 |

| | |
|--|------------|
| 3.9 Carbonyls and Alcohols | 63 |
| 3.10 Carbonyl Reactivity | 68 |
| 3.11 Name Reactions and Synthesis Logic | 69 |
| 3.12 Carbonyls and Alcohols—Key Points | 73 |
| 3.13 Monosaccharides | 74 |
| 3.14 Oligosaccharides and Polysaccharides | 78 |
| 3.15 Chemical Reactions and Tests for Carbohydrates | 78 |
| 3.16 Biological Examples of Carbohydrates | 80 |
| 3.17 Common Nitrogen Compounds | 81 |
| 3.18 Amino Acids and Proteins | 82 |
| 4 Physics | 86 |
| 4.1 Translational Motion | 86 |
| 4.2 Forces and Torque | 86 |
| 4.3 Work and Energy | 88 |
| 4.4 Periodic Motion | 89 |
| 4.5 Fluids and Fluid Statics | 90 |
| 4.6 Viscosity and Poiseuille's Principle | 92 |
| 4.7 Continuity Equation | 93 |
| 4.8 Bernoulli's Equation | 93 |
| 4.9 Turbulence | 93 |
| 4.10 Electrostatics | 94 |
| 4.11 Electromagnetism | 95 |
| 4.12 Electricity and Electric Currents | 95 |
| 4.13 Sound and Doppler Effect | 97 |
| 4.14 Light and Radiation | 98 |
| 4.15 Geometrical Optics | 99 |
| 5 Biochemistry and Molecular Cell Biology | 104 |
| 5.1 Amino Acids and Proteins | 104 |
| 5.2 Carbohydrates & Lipids | 105 |
| 5.3 D- and L- Isomers | 105 |
| 5.4 Monosaccharides, Oligosaccharides, and Polysaccharides | 106 |
| 5.5 Lipids | 107 |
| 5.6 Cellular Reproduction and Chromosomes—Eukaryotic Cells | 108 |
| 5.7 Meiosis | 108 |
| 5.8 Biomembranes and Membrane Transport | 109 |
| 5.9 Nucleus, Nucleolus, & Ribosomes | 109 |
| 5.10 Endoplasmic Reticulum, Golgi Apparatus, & Lysosomes | 109 |
| 5.11 Signal Hypothesis | 110 |
| 5.12 Cytoskeleton | 110 |
| 5.13 Prokaryotic Cells | 110 |
| 5.14 Viruses | 111 |
| 5.15 Metabolic Components—Enzyme Kinetics | 111 |
| 5.16 Metabolic Components—Enzymes and Molecules | 112 |
| 5.17 Metabolic Pathways | 113 |
| 5.18 Genetic Information—Classical Genetics | 119 |
| 5.19 Genetic Loci & Alleles | 119 |
| 5.20 Pedigrees & DNA | 120 |
| 5.21 DNA Synthesis | 120 |
| 5.22 Expression of Genetic Information | 121 |
| 5.23 Mutations | 124 |
| 5.24 Genetic Engineering & Cloning | 125 |

| | |
|---|------------|
| 6 Psychology | 126 |
| 6.1 The Biology Behind Behavior | 126 |
| 6.2 The Human Senses | 129 |
| 6.3 Learning and Memory | 133 |
| 6.4 Emotion, Stress, Motivation, and Addiction | 137 |
| 6.5 Cognition and Consciousness | 141 |
| 6.6 Personality and Attitudes | 146 |
| 6.7 Psychological Disorders | 149 |
| 7 Critical Analysis and Reasoning Skills | 151 |
| 7.1 MCAT Critical Analysis and Reasoning Skills | 151 |
| 7.2 Passage Analysis | 151 |

1 Biology—Physiology

1.1 Types of Tissues

The digestive system is essentially a gastrointestinal canal that is suspended within a body cavity referred to as the **coelom**; this coelom is separated into a **thoracic cavity** (upper, with the lungs and heart) and an **abdominal cavity** (lower, with the liver, stomach, and intestines). There are several types of tissue that you should know about:

- Epithelial Tissues:** Epithelial tissue can be divided into two categories: simple (single layer of cells) and stratified (multiple layers of cells). In these categories, the epithelial cells come in a variety of shapes and sizes: squamous (flat), cuboidal, and columnar. In the gastrointestinal system, simple epithelial cells have projections called microvilli on the luminal (also apical) side. They also have a basal lamina on the basal side that helps hold together epithelial cells using desmosomes. Tight junctions act as permeability barriers and help maintain important gradients for the digestive track. Gap junctions provide a means for water-soluble molecules to pass from the cytoplasm of one cell to the cytoplasm of another cell. If a cell secretes a substance into the lumen by way of a duct, it is referred to as an **exocrine gland**. **Endocrine glands** secrete substances into the blood. Stratified squamous epithelium usually have a protective function. **Fig. 1.1** shows some examples of the types of epithelial tissue.
- Connective Tissues:** Connective tissues include structural, blood cells, mast cells adipose cells, and melanocytes. **Fibroblasts** help secrete the proteins that make up structural connective tissue, such as **collagen** (insoluble, fibrous, strong, flexible), **reticulin** (tin fiber in spleen and lymph nodes), and **elastin** (highly cross-linked fiber associated with organs that require elasticity, like the lungs, skin, and blood vessels). **Cartilage** is secreted by **chondrocytes**, which are a specialized type of fibroblast cell. **Bone** is also a structural connective tissue, made up of collagen, calcium phosphate, and calcium carbonate. Collagen found in bond matrix is secreted by **osteoblasts**, which are specialized fibroblasts. Collagen gives bone flexibility, while the inorganic calcium-based crystals lend rigidity. The central cavity of bone has spongy **bone marrow** where RBCs and WBCs are formed. **Mast cells** can be found in the respiratory tract and GI tract, and they release **histamines** in response to allergic reactions, infections, or injuries. Histamines cause an increase in blood flow to the blood vessels of the affected region. **Adipose cells** are cells that store fat where **Melanocytes** are cells which store pigments.
- Muscle Tissues:** Skeletal muscle is a voluntary muscle, while cardiac and smooth muscle are examples of involuntary muscles.
- Nervous Tissues:** Nerve cells (i.e. **neurons**) are the basic structural unit that make up the nervous system. The major anatomical features are the cell body, **dendrites**, and **axon**. When a neuron becomes excited and receives electrical information, the cell body transmits the signal down the axon in the form of an action potential. The signal causes the release of a chemical neurotransmitter that diffuses across the synaptic cleft and induces an identical action potential in an adjoining neuron, muscle cell, or gland cell. This junction between two such cells is called a **synapse**.

Skin consists of many layers. The **epidermal** region contains **stratified epithelial cells** for protection. Below the epidermis is the **dermis** with a variety of structures. Surrounding the hair follicles are **erector muscles**, which act to straighten the hair shaft. Below the dermis is the **subcutaneous tissue**. This is where one finds adipose deposits.

The inside of the cell is more negatively charged with respect to the outside of the cell. This sets up a voltage that is positive on the outside of the cell. We can calculate the **membrane voltage** (i.e. potential difference) across the cell's membrane using the **Nernst equation**:

$$V_{io} = 2.3 \frac{RT}{ZF} \log \left(\frac{[X]_o}{[X]_i} \right) \quad (1)$$

where V is the voltage in millivolts, i refers to inside, o refers to outside, Z is the ion's valance, and F is Faraday's constant. The inside of a neuron registers a potential difference of about **-80 mV**.

1.2 Action Potentials

Stimulating a nerve causes the membrane to be **depolarized**. If the stimulus is strong enough and the depolarization exceeds the **threshold potential**, then within milliseconds, a burst of Na^+ ions enters and generates an action potential. One millisecond after the burst of Na^+ into the cell, the ion channels that let Na^+ into the cell close and become temporarily

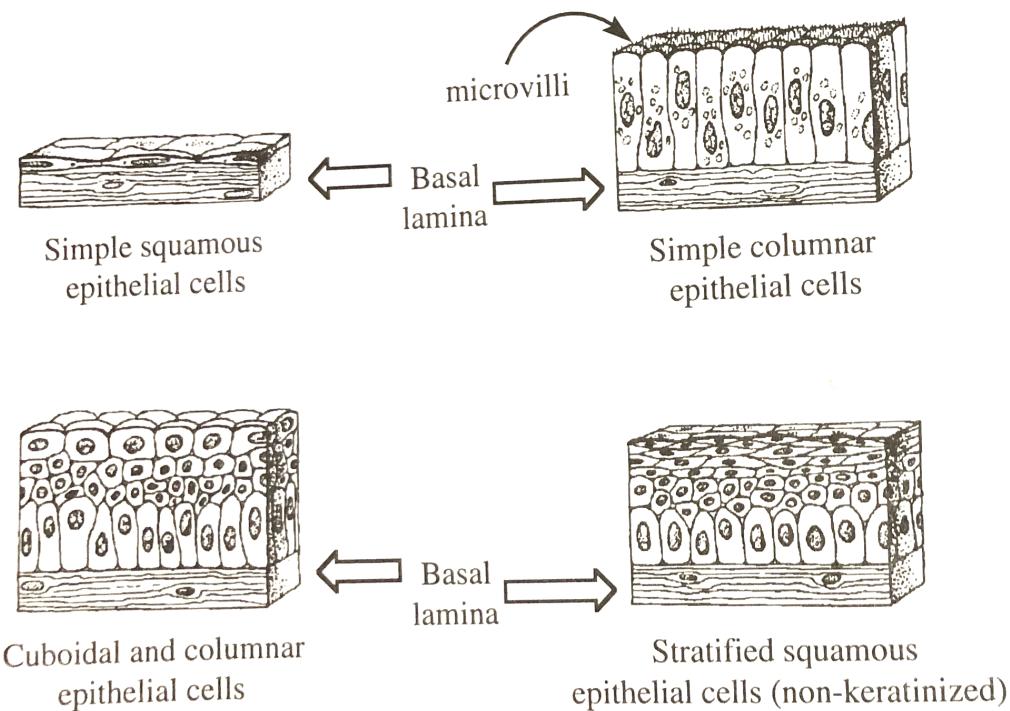
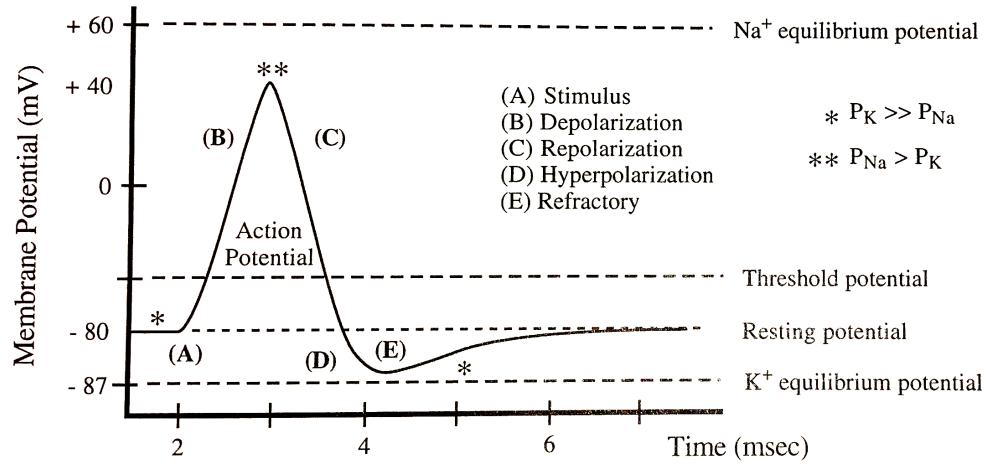


Figure 1: Types of epithelial cells.

inactive—this is referred to as the **refractory period**, which lasts for several milliseconds. During this time a neuron will not be able to generate another action potential, because the Na^+ channels cannot open to allow the cell membrane to depolarize.

Once the Na^+ channels have closed, the K^+ channels begin to open and K^+ begins to leave the cell—this is referred to as repolarization. The overshoot due to the massive outflux of K^+ ions is called **hyperpolarization**. The generation of an action potential is said to be an **all-or-none** phenomenon; an example of a typical action potential is shown in **Fig. 1.2**.

Larger diameter neurons will conduct an action potential further and faster than a small diameter neuron. This means

Figure 2: A typical action potential. P_K is the cell's permeability to K^+ and P_{Na} is the cell's permeability to Na^+ .

that the ability of a neuron to conduct a current depends on the cross-sectional area of that neuron.

Myelinated nerves have the ability to greatly increase the rate at which action potentials are conducted. This is because myelin acts as an electrical insulator and prevents the transfer of ions across the plasma membrane of the axon. This allows

the nerve impulse to ‘jump’ from node to node along the axon, referred to as **saltatory conduction**. Glial cells attach themselves to a section of an unmyelinated axon and begin to rotate around that axon to deposit a myelin sheath. The deposition of myelin along the axon is interrupted by areas where there is no myelin, called the **nodes of Ranvier**. Glial cells in the central nervous system (CNS) are called **oligodendrocytes**, while glial cells in the peripheral nervous system (PNS) are called **Schwann cells**. The CNS is made up of the brain and spinal cord while the **peripheral nervous system** represents the nerves in the periphery.

1.3 The Neuromuscular Junction

At the end of the axon is a specialized area called the **terminal bouton**. Neuromuscular junctions involve a terminal bouton with hundreds of thousands of **synaptic vesicles** that contain the excitatory neurotransmitter **acetylcholine (ACh)**, which is synthesized in the cytosol of the neuron from the molecules **acetyl-CoA** and **choline**. As the action potential reaches the terminal bouton, it triggers the opening of Ca^{2+} channels that generate a transient influx of Ca^{2+} into the terminal region, which causes synaptic vesicles to fuse with the presynaptic membrane and release ACh into the synaptic cleft via exocytosis. Ca^{2+} is pumped out of the cytosol and back into the extracellular fluid. ACh then binds to specific postsynaptic membrane receptors, which conformationally changes the ionophore channel receptor so that it can let ions like Na^+ through. As Na^+ enters the postsynaptic membrane, the muscle fiber begins to depolarize and an action potential is eventually generated. A diagram of the neuromuscular junction is shown in **Fig. 1.3**.

If ACh were to remain in the synaptic cleft, then it would continue to bind to the postsynaptic membrane and action

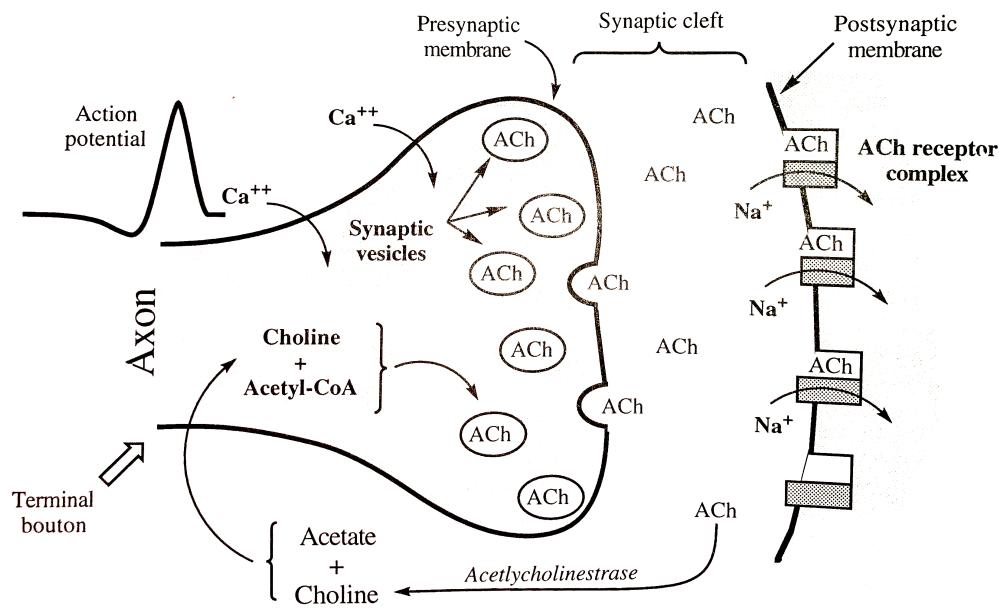


Figure 3: Presynaptic and postsynaptic membranes of a neuromuscular junction.

potentials would continually be generated. This could result in prolonged muscle spasms. However, the enzyme **acetylcholinesterase**, which is bound to the surface of the postsynaptic membrane, hydrolyzes ACh to acetate and choline. Other neurotransmitters include the amino acid **glutamate** and the amino acid derivatives **epinephrine, norepinephrine, dopamine, and serotonin**.

ACh binding to its postsynaptic membrane receptor is referred to as an **excitatory postsynaptic potential (EPSPs)**, since it elicits an excitatory response that generates an action potential. There can also be **inhibitory postsynaptic potentials (IPSPs)**. EPSPs cause an increase in the membrane permeability to Na^+ ions, while IPSPs cause an increase in the permeability of the postsynaptic membrane to K^+ and Cl^- ions.

When two action potentials are coming toward each other in the same neuron, both action potentials will be stopped when they meet. This is because the region in front of each action potential is continuously being depolarized, while the region behind the action potential is in its refractory period for a few milliseconds and cannot immediately be depolarized by the other action potential.

1.4 The Sarcomere

Muscles are multinucleated. They don't divide, but just get larger (i.e. with exercise). Skeletal muscles have **striations** in both the transverse and longitudinal directions. The striations in the longitudinal direction are due to **myofibrils**, which contain the contractile units of the muscle. Each contractile unit, bounded by a structure referred to as the **Z-line**, is referred to as a **sarcomere** and each sarcomere contains two types of contractile proteins. The thin contractile protein is called **actin** and the thick contractile protein is called **myosin**. The myosin thick filament comprises the **A-band**, and the region between two A-bands is the **I-band**. The **H-zone** is that region in the center of the A-band and between the ends of the actin filaments. All these components of the sarcomere are shown in **Fig. 1.4**.

Let's consider the interaction between actin and myosin. When a muscle is in its relaxed state, ATP is bound to the

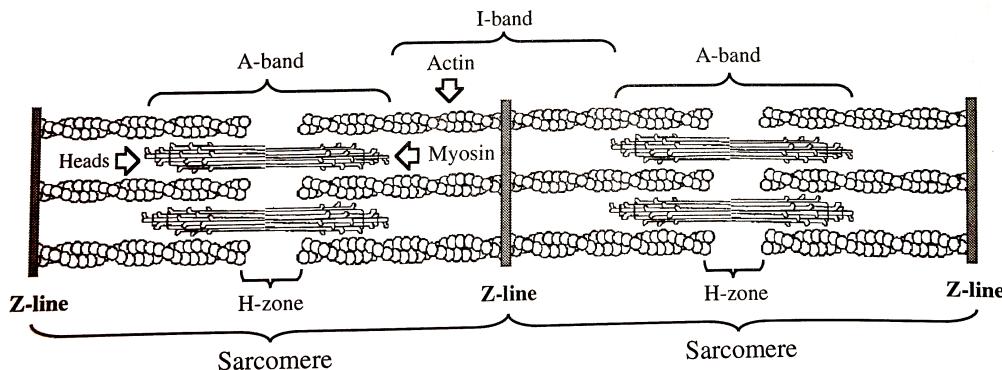


Figure 4: Sarcomere details.

myosin head groups. The myosin head is not bound to the actin filament, because ATP reduces myosin's affinity for actin. The ATP can be hydrolyzed to ADP, and the myosin head now undergoes a corresponding conformational change that is higher in energy and stable. The myosin-ADP-P_i (P_i is the inorganic phosphate hydrolysis product) head complex binds to the actin filament. This step is dependent on Ca²⁺ being present. The interaction between the actin filaments and myosin head groups releases the ADP and P_i from the myosin heads, causing a conformational change in the myosin heads. This step is called the **power stroke**, and the product of this step is referred to as the **rigor complex** or **rigor state**. In order for the myosin heads to dissociate from actin filaments, ATP needs to bind to those myosin heads. This contraction cycle of muscles is shown in **Fig. 1.4**.

Tropomyosin is a protein that runs the length of the grooves of actin's helical structure. When tropomyosin resides in the actin groove, it covers up the binding sites for the myosin heads and prevents those head groups from attaching to the actin filament. **Troponin** is a binding protein that interacts with tropomyosin, actin, and Ca²⁺. When Ca²⁺ is bound to a particular subunit of the troponin complex, it causes tropomyosin to shift its position and expose the myosin head binding sites. Myosin then can bind to the actin filaments, and muscle contraction follows. If Ca²⁺ is not present, then no myosin binding is possible and the muscle is in the **relaxed state**. This interaction is shown in **Fig. 1.4**.

Myofibrils are the basic, rod-like units of a muscle cell. Surrounding each myofibril is a membranous structure called the **sarcoplasmic reticulum**, a modified version of the endoplasmic reticulum. Ca²⁺ is sequestered within this smooth membranous structure. Also surrounding each myofibril is an invagination of the sarcolemma (i.e. the plasma membrane) called the **T-tubule**. Action potentials pass down each T-tubule and stimulate the release of Ca²⁺ from the sarcoplasmic reticulum. Ca²⁺ then bind to the troponin complex, allowing for synchronized muscle contraction. Once contraction has taken place and the nerve impulse has ceased, the Ca²⁺ in the cytosol is pumped back into the sarcoplasmic reticulum by a Ca²⁺-ATPase pump.

The strength of muscle contraction can be varied by (1) the size of the motor unit, which is simply a motor neuron and the muscle fibers that it innervates, (2) the number of available motor units, and (3) the amount of actin and myosin contained within each cell. Smaller motor units means greater precision and lower strength. Less motor units means lower strength.

1.5 Nervous System Components

A grouping of nerve cells is called a **ganglion**. The three basic anatomical division of the vertebrate brain are the *forebrain* (prosencephalon), *midbrain* (mesencephalon), and *hindbrain* (rhombencephalon). Inferior to the brain is the spinal cord,

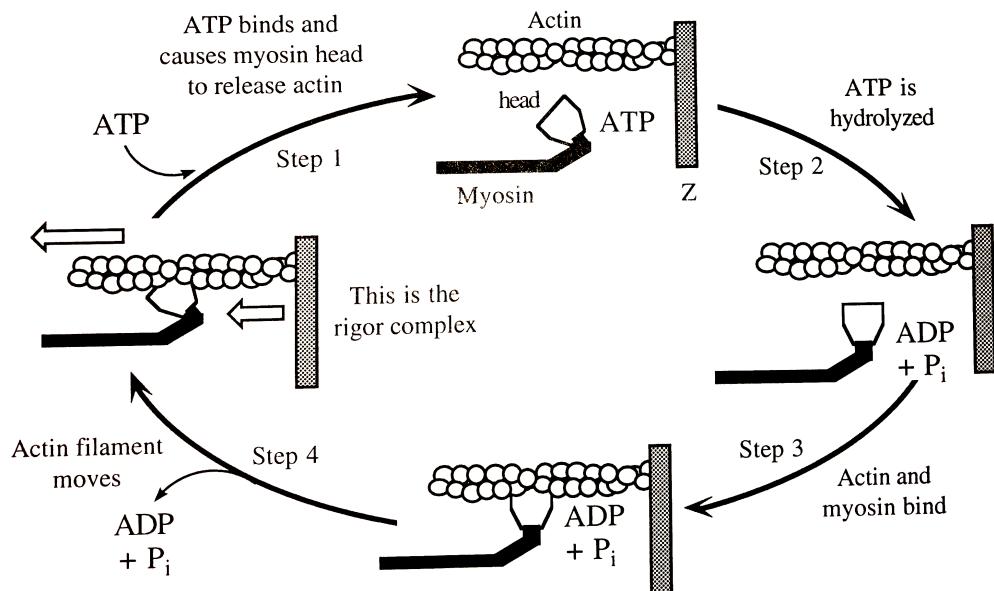


Figure 5: The contraction cycle of actin and myosin.

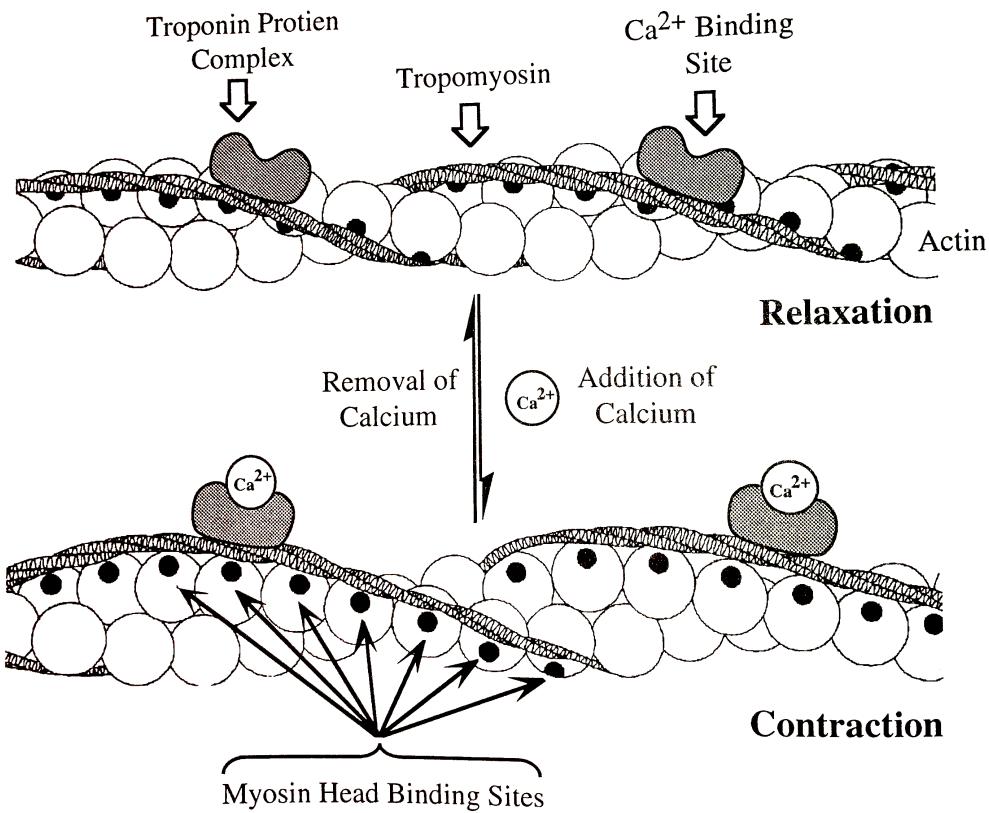


Figure 6: The interaction between troponin, tropomyosin, and calcium in the sarcomere.

from which the nerves extend into the limbs and extremities. If a neuron carries information *into* the spinal cord and brain, that neuron is said to be an **afferent (sensory) neuron**. If a neuron carries the information away from the CNS, that neuron is said to be an **efferent (motor) neuron**.

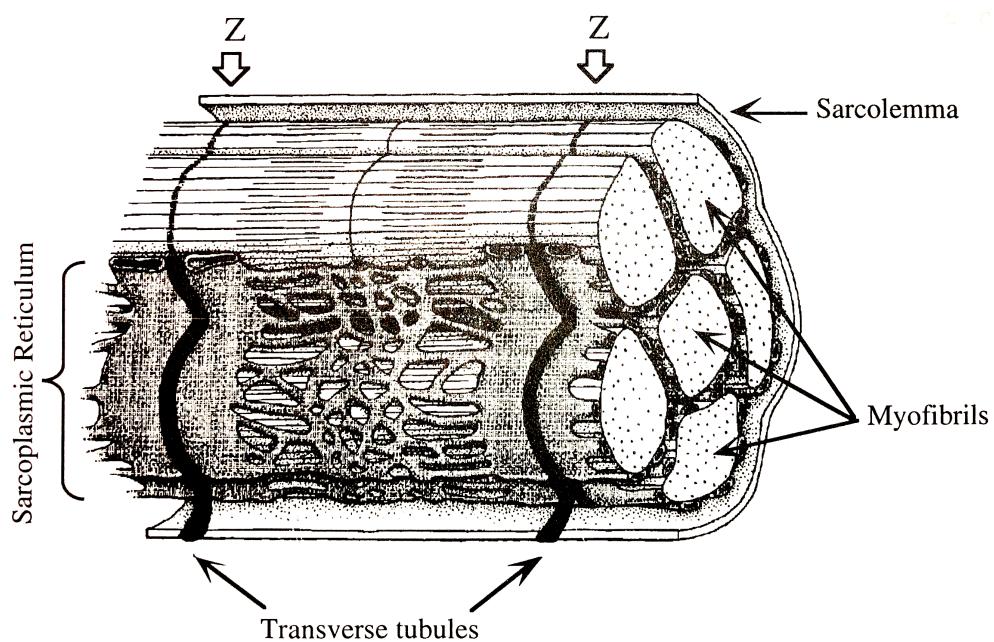


Figure 7: The sarcoplasmic reticulum.

The **prosencephalon** includes the **cerebrum**, **thalamus**, and **hypothalamus**. The cerebrum is divided into the right and left cerebral hemispheres, joined by the **corpus callosum**. The cerebral hemispheres are divided into the **frontal lobes** (associated with movement and personality), **parietal lobes** (associated with touch and stretch sensation), **occipital lobes** (associated with vision), and **temporal lobes** (associated with hearing). The outermost layer of the cerebrum is called the **cerebral cortex**, consisting of **gray matter** (nerve cell bodies and their dendrites) and **white matter** (myelinated axons of the nerve cells). In the spinal cord, the gray matter is more centralized, while the white matter is more peripheral. The anatomical divisions of the vertebrate brain are shown in Fig. 1.5.

The **central sulcus** is a prominent groove that separates the frontal lobes and parietal lobes. Anterior to the sulcus

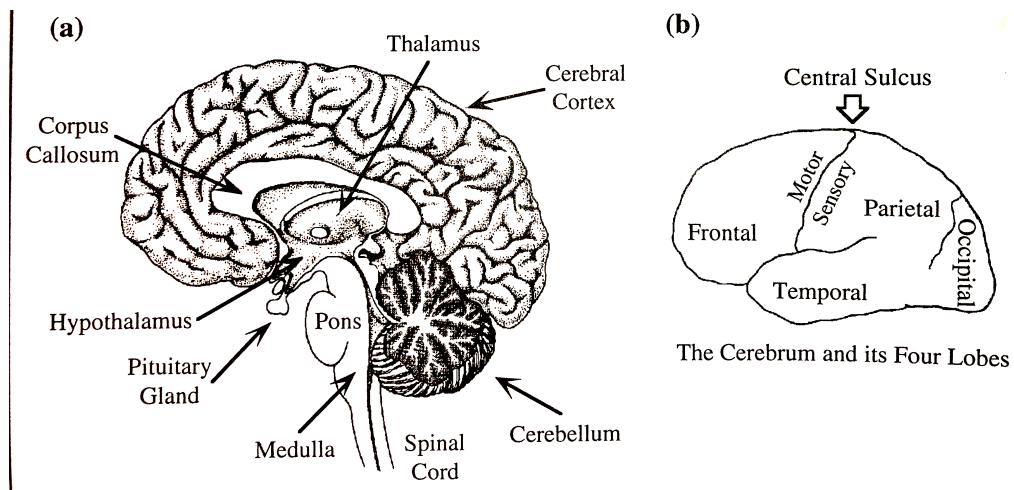


Figure 8: Anatomical divisions of the vertebrate brain.

is the **motor cortex**, which controls the movement of individual muscles. Posterior to this sulcus is the **sensory cortex**, which detects sensations in various parts of the body.

The **thalamus** is a relay station for much of the visual and auditory information from the environment. The **hypotha-**

hypothalamus is concerned with the visceral (subconscious) activities of the body. The **pituitary gland** is the master endocrine gland of the body—it receives information from the hypothalamus and sends out information to regulate different parts of the body.

The **brainstem** contains the **midbrain, cerebellum, pons, medulla**, and the **reticular formation**, which in general coordinate motor and visceral activities. The midbrain detects movement and can direct the head and eyes towards that movement. It can also sense pleasure and pain. The cerebellum is responsible for the bulk of regulation and coordination of muscular activity. The pons and medulla coordinate visceral activities. The reticular formation, which is the core of the brainstem, is essentially an activating system designed to alert the brain. It also inhibits motor and sensory impulses and can induce sleep. Below the medulla is the spinal cord.

1.6 Control of Body Activity and the Neurovisceral Control

A response that makes just one synaptic connection is referred to as a **monosynaptic reflex arc**, while a response that makes at least two synaptic connections is referred to as a **polysynaptic reflex arc**.

The **autonomic nervous system** is part of the **efferent** division of the PNS, which can be further subdivided into the **sympathetic** and the **parasympathetic** systems. The **parasympathetic division** of the autonomic nervous system has nerve fibers which leave from the **sacral** portion of the spinal cord and from the midbrain (mesencephalon) and medulla (part of the rhombencephalon). Parasympathetic nerve impulses tend to increase the rate of digestion and lower the heart rate. The blood pressure is also lowered, and the pupils constrict. In general, the parasympathetic division conserves energy and helps in the restoration of various bodily functions. The parasympathetic division has both **preganglionic** and **postganglionic** nerve fibers. The cell bodies of the preganglionic neurons are found in the sacral region of the spinal cord and in the brainstem. The preganglionic nerve fibers are rather long, while the postganglionic nerve fibers are rather short. Both the preganglionic and the postganglionic nerve fibers in the parasympathetic division release ACh as the neurotransmitter. (Nerve fibers that release ACh as their neurotransmitter are called **cholinergic** nerve fibers.)

The most prominent nerve in the parasympathetic division is the **vagus nerve**, which innervates the heart, lungs, stomach, liver, small intestine, large intestine, and kidneys, among other organs. An overview of the parasympathetic nervous system is shown in **Fig. 1.6**.

The sympathetic division of the autonomic nervous system has nerve fibers branching off from the **thoracic** and **lumbar** regions of the spinal cord. Sympathetic nerve fibers tend to condition the body for a ‘fight-or-flight’ response. The heart rate increases, blood pressure elevates, pupils dilate, and the digestive functions decrease, all as a result of sympathetic nervous innervation. The nerve fibers leaving a synapse in a given ganglion are referred to as **postganglionic** nerve fibers. In the case of the sympathetic division the preganglionic nerve fibers tend to be short, while the postganglionic nerve fibers tend to be longer. The preganglionic fibers release ACh, while the postganglionic fibers release norepinephrine as their neurotransmitters.

An important set of spinal nerves in the sympathetic division are the long preganglionic nerve fibers to the **adrenal medulla**. There are no postganglionic nerve fibers here—when the adrenal medulla is stimulated, both **norepinephrine** and **epinephrine** are released directly into the bloodstream.

The **somatic** nervous system has some notable characteristics:

- Once the nerve fibers leave the CNS, they do not make a synapse until they have reached their effector organ.
- When the synapse is made at the effector organ, the neurotransmitter that is released is **acetylcholine**.
- The somatic nervous system innervates skeletal muscle.
- Innervation of that skeletal muscle leads to excitation of the muscle itself.

In contrast, the **autonomic** nervous system has some notable characteristics:

- Once the nerve fibers leave the CNS, they synapse with a ganglion before they make the final synapse with their effector organ.
- The preganglionic fibers in both the parasympathetic and sympathetic divisions release acetylcholine as the neurotransmitter. The postganglionic fibers in the parasympathetic division release acetylcholine; in the sympathetic division, norepinephrine is released.
- The autonomic nervous system innervates glands, and smooth and cardiac muscle.

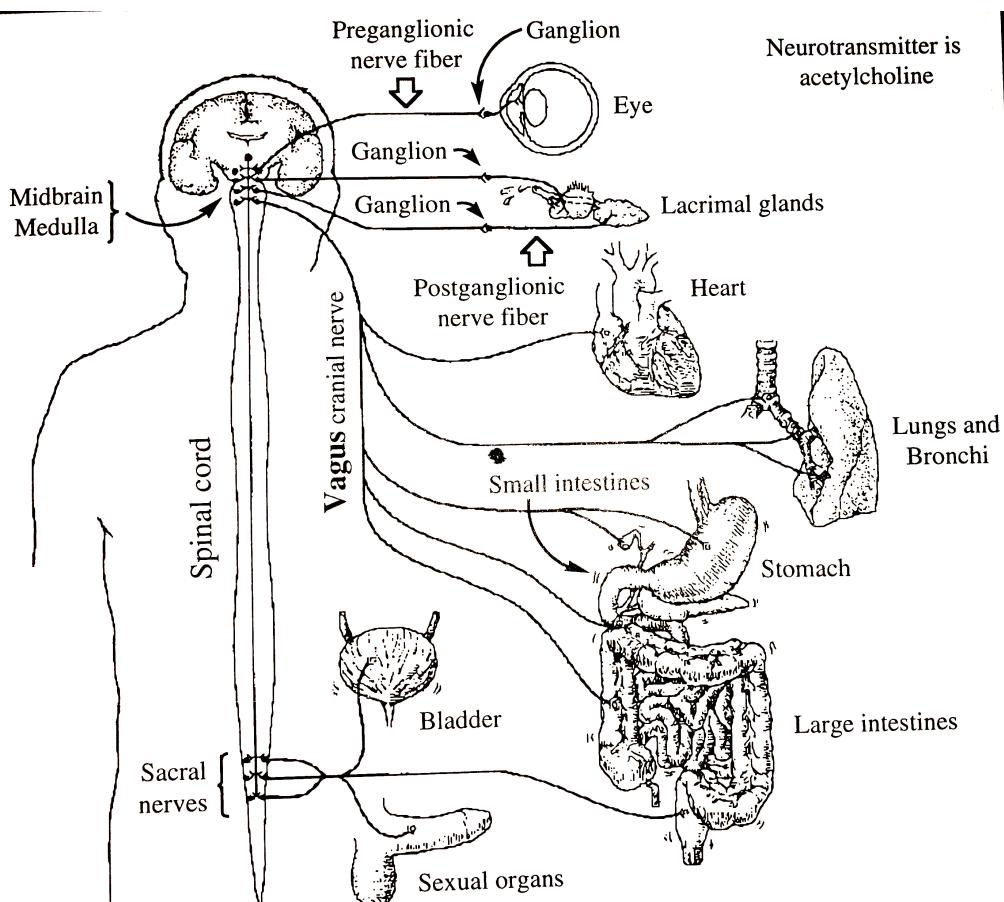


Figure 9: Parasympathetic nerves and the parasympathetic division of the automatic nervous system.

4. The cells innervated by the autonomic nervous system can be either excitatory or inhibitory.

The **adrenal medulla**, a specialized ganglion in the sympathetic division of the autonomic nervous system, is directly stimulated by a preganglionic fiber. This nerve fiber releases ACh, which causes the cells of the adrenal medulla to release epinephrine primarily and also norepinephrine. A review of the CNS and PNS is shown in Fig. 1.6.

1.7 Receptors and Sensory Input

Nociceptors sense pain. In general, the things that are important in propagating signals in the nervous system are the frequency of action potentials, not the magnitude, as the magnitude of action potentials is always basically the same. All receptors essentially function on this basic principle. **Sensory adaptation** is the phenomenon where the body doesn't keep responding to repeated stimuli.

1.8 Cardiovascular Anatomy

The function of the **circulatory system** is to bring nutrients and oxygen to the tissues of the body while simultaneously removing waste products from those same tissues. Because the circulatory system is continually flowing, it also helps to maintain body temperature. Also, the circulatory system can act as a means to transport hormones to various locations within the body. The blood that is pumped from the right heart to the lungs and back to the left heart is called the **pulmonary circulation** while the blood that is pumped from the left heart to the rest of the tissues and back to the right heart is called the **systemic circulation**. Both these circulations lie in series with right another.

Deoxygenated blood passes from the right atrium into the right ventricle, and then is pumped to the **pulmonary artery** and to the lungs where it is oxygenated. The oxygenated blood then returns to the left heart by the **pulmonary veins**.

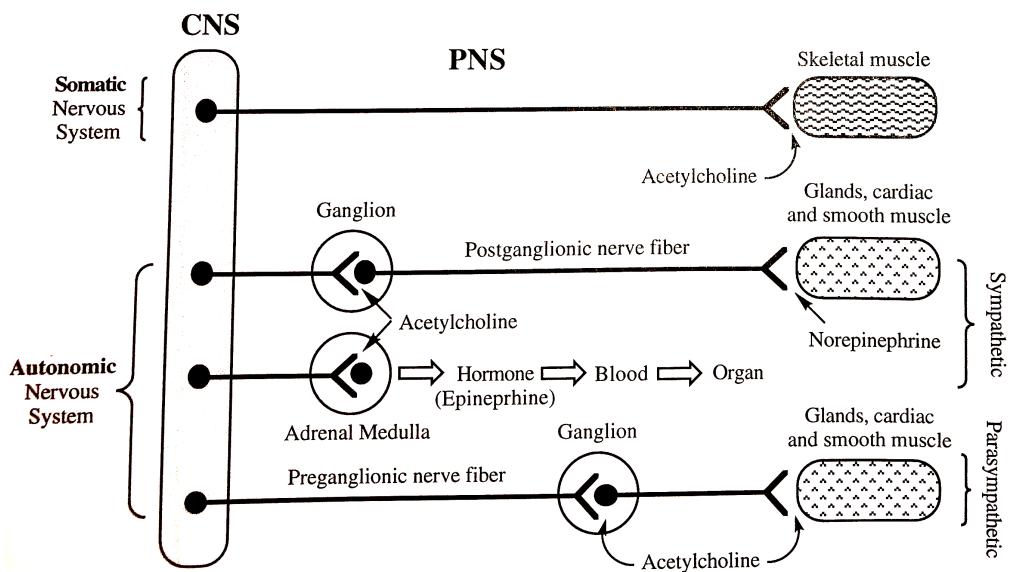


Figure 10: CNS and PNS review.

The blood enters the left atrium and passes into the left ventricle where it is pumped out the **aorta** and to the branching arteries, arterioles, and capillaries. It is at the level of the capillaries that the blood exchanges nutrients and oxygen for waste products created by metabolism. Deoxygenated blood passes from the capillaries to **veins** and then to larger veins, and eventually to the **superior** and **inferior vena cava**, which enter the right atrium. The general anatomy of the heart is shown in **Fig. 1.8**.

As the left ventricle contracts (**systole**), it propels blood out the aorta and into the arteries with a pressure of about

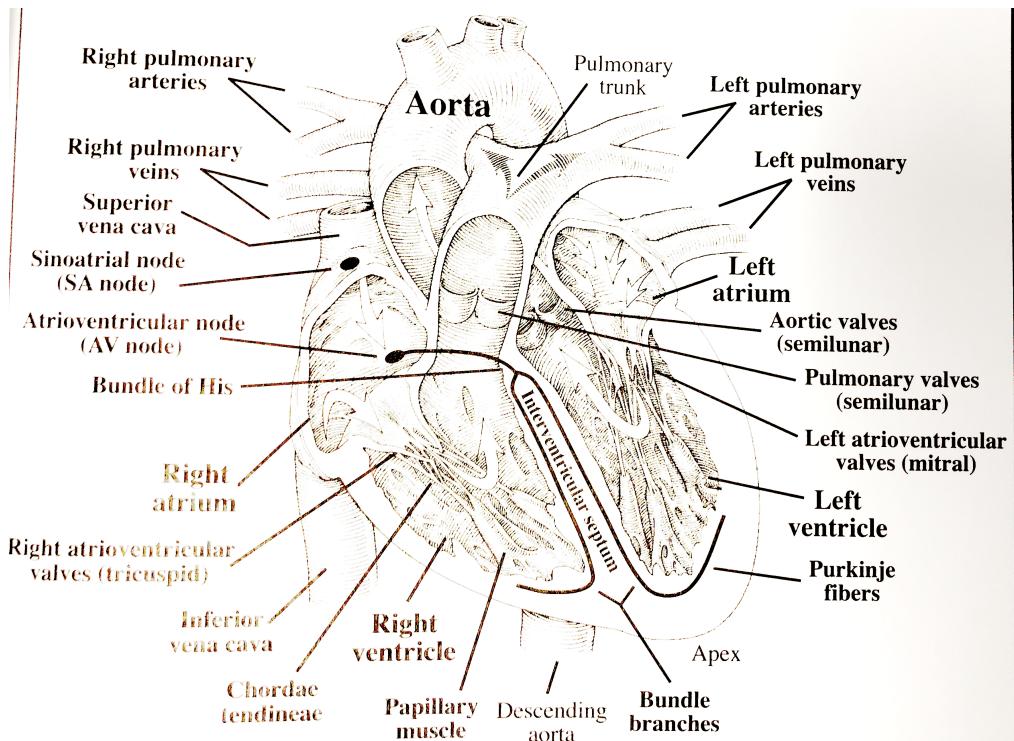


Figure 11: Anatomical landmarks of the heart.

120 mmHg. After contraction takes place and ventricles begin to relax (**diastole**) and fill with blood, the pressure in the arteries is about 80 mmHg. Since blood is under a lot of pressure, the arteries have to be durable. They have thick walls composed of **smooth muscle** and **connective tissue** that contain both **collagenous** and **elastic** fibers.

The lumen of blood vessels are lined with epithelial cells called **endothelial cells**. Damage to these cells by the pulsating arterial pressure or even by abrasive substances in the blood can lead to **atherosclerosis**. Once these cells are damaged, cholesterol can build up at the site of the lesion and a plaque will form. During the later stages, the arteries become ‘hardened’ from the layers of deposition, and this is referred to as **arteriosclerosis**.

Regulation of the circulatory system is controlled by the sympathetic and parasympathetic divisions of the autonomic nervous system. The sympathetic division is the more important of the two. Besides nervous control of blood flow, there is also humoral control from the action of ions or hormones and local control at the level of the individual tissues from various metabolites.

The wall of capillaries are composed of a unicellular layer of endothelial cells. Surrounding these cells is a basement membrane. However, there is no connective tissue or smooth muscle. The capillary itself is just large enough for a RBC to squeeze through. At the entrance to the capillary bed is a **precapillary sphincter** composed of smooth muscle which helps to regulate the flow of blood to the area.

Valves are very important in maintaining the directionality of blood flow. There are atrioventricular valves between the atria and the ventricles, which close as the ventricles are contracting, and also the pulmonary valve and aortic valves that close when blood flows out of the pulmonary artery and the aorta, respectively. The atrioventricular valves do not invert because of tendinous cords called **chordae tendineae** that hold them in place.

Located near the junction of the superior vena cava and the right atrium is the **sinoatrial node** (SA node), which is the pacemaker of the heart. The SA node is the point of origin for the electrical impulse that propagates through the rest of the heart. Located in the lower portion of the right atrium and near the right ventricle is the **atrioventricular node** (AV node). Impulses from the SA node also spread to the AV node and then from the AV node through a collection of fibers called the **bundle of His**. Branches of the bundle of His surround the ventricles, and when this bundle receives an impulse, it causes the ventricles to contract and eject blood to the pulmonary and systemic systems.

Sympathetic nerves and norepinephrine can be released at the SA node to increase the rate of the heart. **Cardiac output** is the amount of blood that is pumped per minute by each of the two individual ventricles of the heart:

$$\text{Cardiac Output} = (\text{Stroke Volume}) (\text{Heart Rate}) \quad (2)$$

where stroke volume is measured in liters per beat and heart rate is measured in beats per minute. **Poiseuille's Law** describes the flow rate of a fluid in blood vessels:

$$\text{Flow Rate} = \Delta P \frac{\pi R^4}{8\eta L} = (P_1 - P_2) \frac{\pi R^4}{8\eta L} \quad (3)$$

where ΔP is the pressure drop, R is the radius of the blood vessel, η is the coefficient of viscosity, and L is the length of the tube. The actual formula for Poiseuille's Law is not so important—rather, understanding how flow rate changes with each of these physical variables is much more important. Meanwhile, **Fick's Law** describes the net flux/net rate of diffusion J in mol/sec:

$$J = -DA \frac{\Delta C}{\Delta x} \quad (4)$$

D is the diffusion coefficient (a proportionality constant), A is the area of the plane of interest, and $\Delta C/\Delta x$ is the concentration gradient across that plane. The minus sign indicates the direction of the flux or diffusion.

Osmolarity is basically normality. For example, if we had a 1M solution of glucose, it would have a concentration of 1 osmol per liter. If we have a 1M solution of sodium chloride, this has a concentration of 2 osmols per liter (one from the Na^+ ion and one from the Cl^- ion).

1.9 Lymphatic System

The **lymphatic system** lies parallel to the systemic and pulmonary circulations. It collects the excess fluid that leaks into the interstitial space from the capillaries and returns it by way of the vena cava back to the circulatory system. Lymph nodes located along the lymphatic system help to filter out foreign particles that could potentially lead to disease. If the lymph flow through the lymphatic system were blocked, **edema** would result. This is simply an increase in the interstitial fluid.

1.10 Blood Clotting

Blood clotting occurs via a *cascade process*. As a general overview, **thrombin** is a serine protease involved in blood clotting. There are two pathways in blood clotting—the **intrinsic route** (due to contact with some abnormal surface) and the **extrinsic route** (due to trauma to the tissue). In blood clotting there are more than 15 different factors involved, and about 8 or so of them are serine proteases.

In **Fig. 1.10**, we show the intrinsic and extrinsic pathways. The subscript ‘a’ means that we are dealing with the *active* form of the molecule. Factor IX will be converted to Factor IX_a by some **intrinsic factor**. Factor IX_a will be the trigger which will convert Factor X to Factor X_a. From the **extrinsic** portion of this scheme, we start with some tissue factor which converts Factor VII to Factor VII_a. Factor VII_a can also bring about the conversion of Factor X to Factor X_a. It is then Factor X_a (and also an additional factor, Factor V_a) that is involved in the conversion of **prothrombin** (Factor II) to **thrombin** (Factor II_a). Thrombin converts **fibrinogen** (Factor I) to **fibrin** (Factor I_a). These fibrin fibers are then cross-linked by Factor XIII_a, which is an enzyme called **transglutaminase** to form the mature cross-linked fibrin clot. **Vitamin K** is a

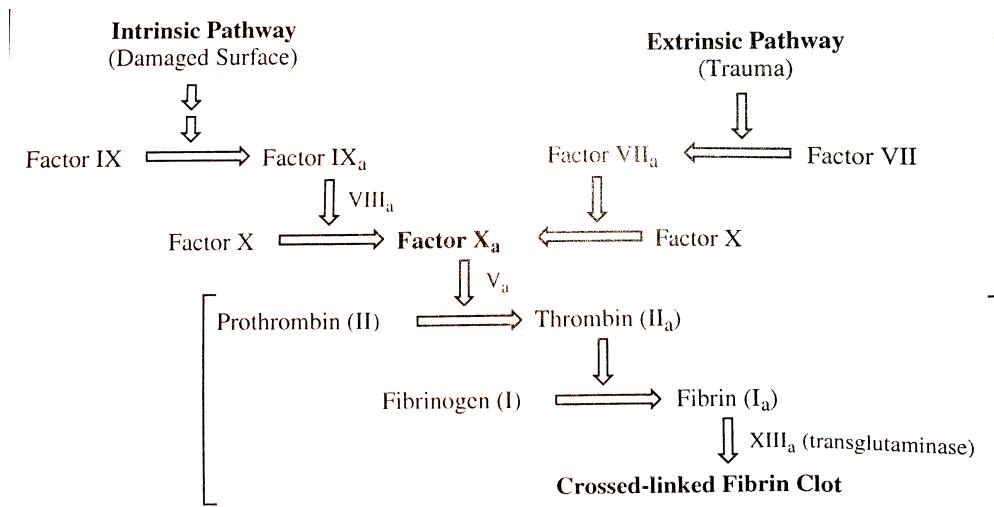


Figure 12: The intrinsic and extrinsic pathways of blood clotting.

fat-soluble vitamin, and it is important for blood clotting. This is because prothrombin exists in an even earlier form called *preprothrombin*, which has certain Glu residues that are carboxylated by a carboxylase enzyme. This carboxylase enzyme has an absolute requirement for Vitamin K. Therefore, in the presence of Vitamin K, HCO_3^- , and the carboxylase enzyme, we can convert preprothrombin to prothrombin. This structure now has great affinity for divalent ions such as Ca^{2+} , and is therefore a good chelating agent. Therefore, the essentials for the clotting process to take place is the (a) platelet membrane, (b) enzymes, (c) Ca^{2+} , (d) an auxiliary factor, Factor V_a, and (e) the substrate prothrombin.

Thrombin can now act as a proteolytic enzyme that converts **fibrinogen** to **fibrin**. Fibrinogen is a large soluble protein. Its solubility is due to an excess of negatively charged amino acids (i.e. Glu, Asp, Tyr-SO₄), particularly in the central domain of the molecule. The net charge in the central domain of the fibrinogen is -8, while the net charge at the terminal ends is -4. In general, if very large molecules have a net charge of zero, they will tend to come together. However, if we have an excess of charge (as is the case with fibrinogen), the molecules will repel one another. Thus, in order to convert the soluble fibrinogen to the insoluble fibrin, we can remove the portion of the molecule contributing all the negative charge. Indeed, when fibrinogen is converted to fibrin, we find that 4 Arg-Gly bonds are broken in the central domain, releasing four peptides containing an excess of negative charge. Once these ‘fibrinopeptides’ are released, the overall net charge in the central domain now becomes +5 (originally -8 in fibrinogen). The fibrin monomer that is now formed has the ability to interact with other fibrin monomers through electrostatic interactions between the terminal (negatively charged) and central (positively charged) regions of the polypeptide—the resulting aggregation is called a **fibrin (soft) clot**.

To make the soft clot into a hard clot, the enzyme transglutaminase (Factor XIII_a) cross-links the monomers. Once the damaged area has been repaired, a serine protease called **plasmin** hydrolyzes specific regions in the fibrin clot in order to dissolve it into smaller peptide fragments (to remove the clot). **Tissue plasminogen activator (TPA)** converts **plasminogen** into this active protease.

1.11 The Lungs

When we breathe air, it first enters through the nose/mouth, passes from the oral cavity to the **pharynx**, into the **larynx**, and then down the **trachea**. At the end of the trachea, the air passes into two tubular passageways called **bronchi**. One bronchi enters into each lung and continues to divide into smaller passageways called **bronchioles**, eventually ending in the functional units of the lungs called the **alveoli**. Each alveolus consists of a single layer of epithelial cells juxtaposed to a very thin basement membrane. Surrounding each single alveolus is a capillary network.

At equilibrium, the number of O₂ molecules dissolving in water will equal the number of O₂ molecules leaving the water. This means the partial pressure of oxygen in the gas phase (air) is equal to the partial pressure of oxygen in the liquid phase (water):

$$(P_{O_2})_{\text{gas}} = (P_{O_2})_{\text{liquid}} \quad (5)$$

The **epithelial cells** that line the lumen of the passageways to the end of the bronchioles have **cilia** which are continually beating towards the pharynx. Scattered among these epithelial cells are glands that secrete **mucus**, which is important to trap and foreign matter and get rid of it as the mucus is moved upward towards the oral cavity. The upper passageways of the respiratory tract maintain their opening by means of **cartilage rings**. Smooth muscle is found in almost all areas of the respiratory tract where there's no cartilage. **Asthma** is caused by allergic hypersensitivity to airborne antigens that have entered the respiratory tract. The direct result is to cause the mast cells within the bronchioles to release a number of different substances that cause the smooth muscle surrounding bronchioles to spasm and constrict.

The lungs are encased in the pleural membrane. The **visceral pleura** covers the lungs, while the **parietal pleura** adheres to the diaphragm and chest wall. Between the visceral pleural and parietal pleura is the **intrapleural space** which contains a watery fluid. During inspiration, the muscles of the diaphragm contract and pull the diaphragm down, enlarging the thoracic cage and expanding the lungs. This creates a subatmospheric pressure in the alveoli, which causes air to rush down its pressure gradient from the atmosphere into the lungs. During expiration, the thoracic cage and lungs return to normal, and air in the alveoli is compressed and forced out through the passageways of the respiratory tract. If the lungs were to become separated from the visceral pleura, they would collapse, causing **pneumothorax**. This is because they have no anatomical structures to maintain rigidity.

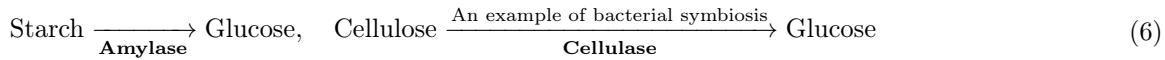
As deoxygenated blood comes back from the tissues, it enters the right atrium via the superior and inferior vena cava and passes into the right ventricle, where it is pumped to the lungs via the pulmonary artery. In the alveoli, both O₂ and CO₂ diffuse down their respective concentration gradients from the capillaries to the alveoli. This means that the oxygen enters the blood and carbon dioxide leaves the blood. The blood then becomes oxygenated, comes back to the heart via the pulmonary vein, and passes from the left atrium to the left ventricle and then pumped to the tissues via the aorta. Oxygen can travel/dissolve in the blood itself, but it's pretty insoluble in water/blood. Rather, most oxygen is transported by being bound to a transport protein in erythrocytes called **hemoglobin (Hb)**. Hb has four polypeptide units, each of which has a **heme** prosthetic group with an iron atom in the center in the ferrous Fe²⁺ state. Thus, each hemoglobin molecule can take up to four O₂ molecules.

Hb is dependent on pH and temperature—if we were to decrease the pH of the blood or increase the temperature, then less hemoglobin will be saturated with oxygen. This is what happens when you exercise. Similarly, this shift in the oxygen saturation curve happens when a byproduct of glycolysis, **2,3-bisphosphoglycerate (2,3-BPG)**, binds to hemoglobin.

CO₂ can be carried in the blood by either (1) dissolving in the plasma and RBCs, (2) binding to a specific site on the hemoglobin molecule, or (3) in the form of bicarbonate ions (HCO₃⁻). Most of the CO₂ is carried in the blood in the form of bicarbonate ions.

1.12 The Gastrointestinal Tract

The gastrointestinal (GI) system includes the mouth and associated salivary glands, esophagus, stomach, small intestine, large intestine, and certain aspects of the liver and pancreas. Starch, characteristic of plant cells, and glycogen, characteristic of animal cells, are hydrolyzed within the digestive tract by **amylases**. Upon hydrolysis, both polymers release glucose. **Cellulose** is found within the cell walls of plants; the enzyme **cellulase** can hydrolyze cellulose into its constituent glucose residues.



Proteases can hydrolyze proteins to their constituent amino acid residues. Of the 20 canonical amino acids, we need to obtain 9 of them in our diet, which are thus referred to as **essential amino acids**. The rest we can synthesize ourselves. Fat

cells (adipocytes) store triacylglycerols (i.e. fats) that are hydrolyzed into fatty acids and glycerol by the enzyme **lipase**.

In the small intestine, the innermost layer is a convoluted layer of epithelial cells, with endocrine cells and ducts from external exocrine glands like the pancreas and liver scattered throughout that can release hormones into the blood and influence other cells in the GI tract. Both the parasympathetic and sympathetic nerves of the autonomic nervous system innervate the GI system; the more important nerves stem from the **parasympathetic** system. Recall that the parasympathetic nerves are more active during times of relaxation and digestion, while the sympathetic nerves are more active during times of ‘fight or flight.’

As food is swallowed and passed into the pharynx, access to the nasal cavity is closed; a flap of tissue called the **epiglottis** covers the opening to the **larynx** and prevents food from entering into this passageway to protect the airway. The swallowing reflex is controlled by centers in the medulla.

As food boluses pass through, the smooth muscles surrounding the epithelial cells contract in peristaltic waves, and this is referred to as **peristalsis**, controlled by the action of the parasympathetic division and by the action of hormones.

Once the food reaches the stomach, a *sphincter* (circular muscle) called the **gastroesophageal sphincter** contracts and prevents regurgitation of the food back into the esophagus. If this sphincter were not closed off, stomach acid would enter the esophagus and irritate the nerve endings in the smooth muscle, causing a burning sensation referred to as heartburn. In the stomach, there are four major types of secretion. **Mucus** is secreted by **surface cells** and acts to protect the lining of the stomach and lubricate the food. **Gastrin**, located in the endocrine cells in the lower portion of the stomach, is secreted in response to protein entering into the stomach. This hormone stimulates the secretion of HCl and pepsinogen. **HCl** at a pH of about 1 is secreted by **parietal cells**, while **pepsinogen** is secreted by the **chief cells**. Parietal cells and chief cells are located in the **gastric pits** that line the epithelium of the stomach. Pepsinogen is the inactive form of the peptidase enzyme **pepsin** that becomes activated by HCl and also by a positive feedback loop and auto-activation once pepsinogen reaches the stomach.

In addition to secreting HCl, the parietal cells also secrete a glycoprotein called **intrinsic factor**. This glycoprotein is important because it complexes with vitamin B₁₂ and is then absorbed by the intestinal epithelial cells and transported by the bloodstream. Vitamin B₁₂ is important in erythrocyte formation. If too much acid is secreted into the stomach, **ulcers** can occur in the stomach and small intestine. Ulcers are erosions of the walls of these organs, and if they are extensive, they can cause bleeding. **Histamine** is a powerful stimulant that causes HCl to be released into the lumen of the stomach, and so compounds like **cimetidine** inhibit the binding of histamine to its receptor on parietal cells. This effectively reduces the amount of HCl secreted in the lumen of the stomach.

Roughly 90% of digestion and absorption that takes place in the GI tract occurs in the small intestine, which acts to neutralize the acid which has been secreted by the stomach and to further digest and absorb food particles. Distension of the small intestine also causes the hormone **cholecystokinin (CCK)** to be released from the intestinal mucosa. CCK diffuses through the bloodstream to the pancreas where it causes the pancreas to release digestive enzymes. The hormone **secretin** is released from the small intestine in response to the entering chyme from the stomach. Secretin is also absorbed by the blood and is transported to the pancreas where it causes the release of bicarbonate ion and other fluids.

The **pancreas** not only has endocrine cells that secrete insulin and glucagon, but also contains secreting structures called **acini** that secrete a fluid into the small intestine which has a high **bicarbonate** content that is rather alkaline, which reacts with the stomach acid to produce carbonic acid, which then decomposes into carbon dioxide and water.

The liver has many functions, one of which is to synthesize **bile** that is concentrated and stored in the **gallbladder**. Bile contains a major pigment called **bilirubin**, which is the breakdown product of hemoglobin. It also contains **bile salts** which are important in the digestion and absorption of fats. When bile is released from the gallbladder, it passes down a duct that joins with the pancreatic duct, through a constriction called the **sphincter of Oddi**, and empties into the small intestine. Bile acts to emulsify fats, which basically means just decreasing their surface tension in order to break them up into smaller sizes, making it easier to absorb. The presence of fats in the small intestine releases CCK which then acts on the gallbladder and causes contraction, and on the sphincter of Oddi and causes relaxation, so the bile can pass into the lumen of the small intestine.

Remember that the Na⁺-glucose symport is on the apical side of the epithelial cells in the small intestine (the side facing the lumen), while the Na⁺-K⁺ antiport pump is on the basolateral side facing the blood. There's also a glucose passive transporter on the basolateral side to allow glucose to naturally enter the blood according to its concentration gradient. This is shown in Fig. 1.12.

Recall that fats are degraded to fatty acids and glycerol by the lipase enzyme. These compounds can diffuse into the intestinal epithelial cells where they are resynthesized into triglycerides (fats) and aggregate into structures called **chylomicrons**. These aggregates are released at the basolateral membrane and into the extracellular space, and enter the lymph and are transported to the veins and eventually the tissues.

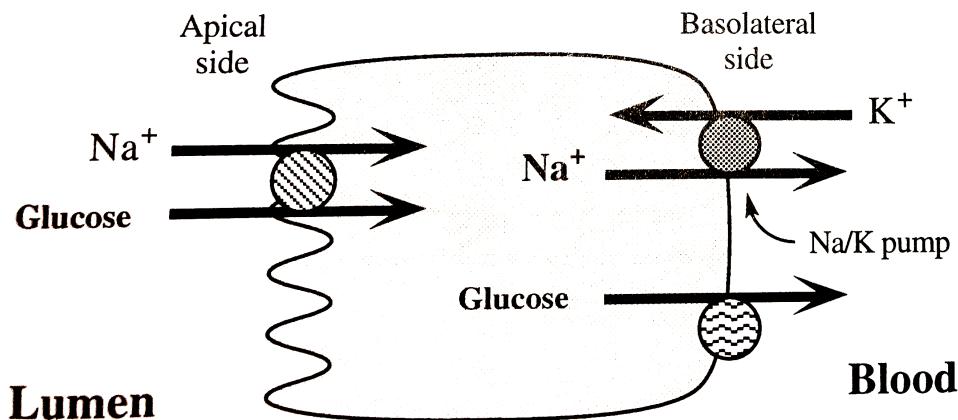


Figure 13: Transport systems.

The large intestine absorbs most of the water and ions that are left in the chyme as it passes from the small intestine. As the sodium and chloride ions are absorbed, an osmotic gradient is established that allows for the absorption of water. What is not absorbed is passed out of the body in the feces.

1.13 The Kidney

Water can be lost from the **integument** (i.e. the skin) by evaporation, the GI tract, the lungs, and the urinary system. Recall that osmolarity refers to the total solute concentration of a solution. Remember, the higher the osmolarity of a given solution, the lower will be the concentration of water in that solution.

If an organism can change the internal ionic concentration of its fluids to meet that of the surrounding environment are referred to as **osmoconformers**; many marine invertebrates are osmoconformers. However, most vertebrates do not change the internal ionic concentration of their body fluids to meet that of a surrounding environment. Organisms like this are referred to as **osmoregulators**.

Kidneys have three main functions: filtration, reabsorption, and excretion.

- Filtration:** The functional unit in the kidney is the **nephron** which consists of a **glomerulus**, **Bowman's capsule**, and a **tubular** system. The glomerulus is a collection of capillaries that receives blood from an artery terminating in the renal system. Blood is pumped into the glomerulus by the hydrostatic pressure of the heart, and that pressure forces the blood through the capillary walls and into Bowman's capsule. The cell-free ultrafiltrate found in Bowman's capsule lacks many of the plasma proteins found in the blood. Blood **plasma** is simply a solution which is about 90% water that is composed of organic (i.e. proteins, sugars, amino acids, etc.) and inorganic substances (i.e. various ions like sodium and chloride). There's also erythrocytes, lymphocytes, and platelets. The **filtrate** in Bowman's capsule is essentially the plasma minus the large MW proteins.
- Reabsorption:** The kidney also reabsorbs important organic and inorganic compounds from the filtrate in Bowman's capsule. Reabsorption occurs through many of the epithelial cells which line the tubular lumen of the nephron. The **cilia** of these epithelial cells move in such a way as to help propel the filtrate through these renal tubes. Glucose, small proteins, amino acids, salts, bicarbonate ions, and water are reabsorbed along this tubular system and transported back into the blood. This epithelial cells also have the ability to secrete protons, potassium, urea, uric acid, and ammonia.
- Excretion:** The kidney also helps excrete waste, salts, and excess water.

Blood from the descending aorta enters into the **renal artery** and eventually into the **glomeruli** of the nephrons. Blood leaves the kidney by way of the **renal vein** which itself empties into the inferior vena cava. Leaving each kidney is a **ureter** which transports the urine to the **bladder**. Urine exits the bladder by way of the **urethra**. Each kidney contains more than a million nephrons. The glomerulus and Bowman's capsule of each nephron is located in the **cortex** and gives this portion of the kidney a **granular** appearance. In contrast, the **striated** appearance of the **medulla** is primarily due to a portion of the system of the nephron called the **loop of Henle** and the **collecting duct**. The collecting ducts, which collect the

urine, empty into the renal pelvis of the kidney and then into the ureter.

The functional unit of the kidney is the **nephron**. As the **renal artery** branches into smaller division and enters the kidney, it becomes the **afferent arteriole**, which then enters into the **Bowman's capsule** and forms the capillary bed called the **glomerulus**. An **efferent arteriole** leaves the glomerulus and forms a capillary network which surround the renal tubules. The blood that leaves this capillary network does so by a **venule** which later empties into the **renal vein** that leaves the kidney. Extending from Bowman's capsule is a long tubular structure divided into the **proximal convoluted tubule (PCT)**, **loop of Henle**, **distal convoluted tubule (DCT)**, and the **collecting duct**. Many different nephrons can attach to a single collecting duct. The anatomy of the nephron is shown in Fig. 1.13.

The PCT is the *obligatory* section of the nephron because roughly 65% of all reabsorption and secretion occurs here.

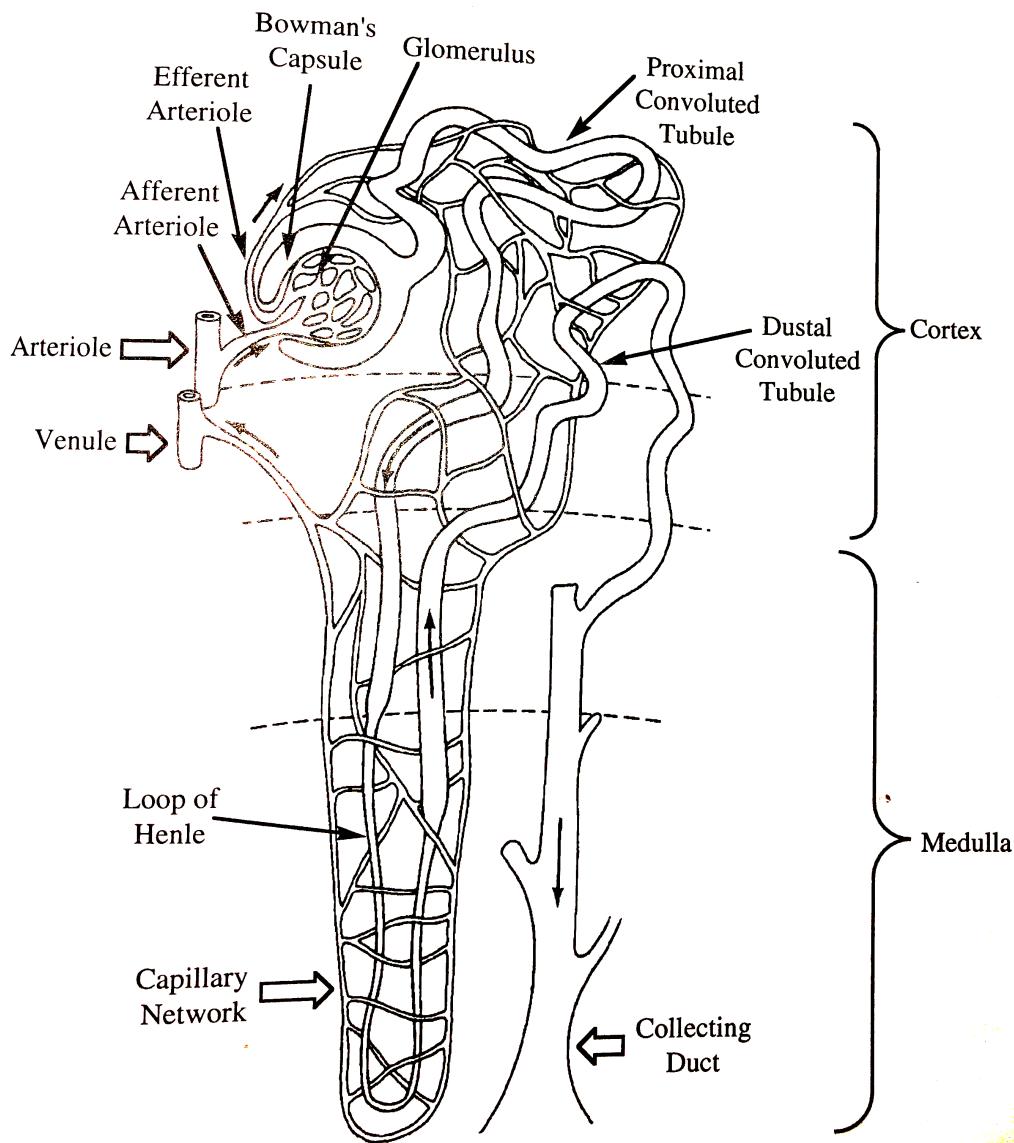


Figure 14: The nephron and its related capillary system.

Little regulation occurs in the PCT. Glucose, small MW proteins, amino acids, and vitamins are *completely* reabsorbed in the PCT, and roughly 80% of Na^+ , Cl^- , and water are reabsorbed in the PCT. As the PCT begins to descend from the cortex into the medulla, it forms the loop of Henle, which has a thin portion and thick portion. The descending thin portion is very permeable to water but only relatively permeable to ions and molecules. The ascending thin portion is the reverse. As the loop of Henle ascends it becomes thicker. The epithelial cells in this region actively transport ions like Na^+ and K^+ from

the lumen of the loop into the interstitial fluid. However, this region is impermeable to urea and water. The filtrate from the thick portion of the loop of Henle passes into the DCT, which is also impermeable to urea and water but rather permeable to ions. The filtrate becomes even more dilute. The DCT epithelial cells and the collecting duct cells located in the cortical region of the kidney are quite sensitive to **aldosterone** secreted by the adrenal glands. An increase in aldosterone causes sodium to be reabsorbed by the epithelial cells. These cells also respond to **antidiuretic hormone (ADH)**, produced by the hypothalamus and released by the posterior pituitary. If the concentration of ADH increases, then water will be reabsorbed from the epithelial cells in the collecting duct and passed to the interstitial space. The cells of the collecting duct can also secrete hydrogen ions (as can the cells of the PCT and DCT).

The osmolarity of the filtrate in Bowman's capsule is essentially the same osmolarity in the plasma.

1.14 Homeostatic Mechanisms

Suppose we want to maintain the levels of sodium in the body. Certain cells in the cortex of the adrenal gland are sensitive to the levels of sodium in the blood. When the concentration of Na^+ decreases in the blood, these cells release the hormone **aldosterone**, which acts at the level of the **DCT** (distal convoluted tubule) and the **collecting ducts** and stimulates the epithelial cells in those regions to reabsorb sodium. As a result, the blood concentration of sodium begins to rise. This acts as a signal and feeds back on the cortical cells of the adrenal gland, telling it to reduce the secretion of aldosterone.

Suppose we want to maintain the proper levels of water in the body. If there's a *decrease* in the plasma volume of the body, there will be a lower blood pressure (detected by baroreceptors) but a higher osmotic pressure (detected by osmoreceptors). These receptors would stimulate specific cells in the hypothalamus to synthesize and transmit **ADH** (anti-diuretic hormone) to the posterior pituitary where it is released into the blood. ADH stimulates the epithelial cells in the latter portion of the **DCT** and collecting ducts to absorb water. This leads to less water excretion and a higher plasma volume. If you were to drink too much water, then the opposite process would happen and you would excrete copious amounts of urine.

We also should talk about nitrogenous wastes. Proteins and nucleic acids are the two primary metabolic sources of nitrogenous wastes, and can be removed as ammonia, urea, or uric acid. Ammonia is toxic and soluble, urea is less-toxic and soluble, and uric acid is toxic and insoluble. Fish metabolize glutamine into ammonia, which can be easily carried away with the passing water. Urea is excreted by most mammals, amphibians, and some reptiles and birds. The ammonia that is produced from the metabolism from the amino acids Gly, Asp, and Glu can be converted to uric acid, which is the primary excretory product of birds and reptiles.

The normal pH of the body is 7.4. Anything below 7.35 or above 7.45 could be fatal if left for an extended period of time. Blood pH is regulated by a **buffering system** that involves bicarbonate: $\text{H}^+ + \text{HCO}_3^- \rightleftharpoons \text{H}_2\text{CO}_3 \rightleftharpoons \text{CO}_2 + \text{H}_2\text{O}$.

Recall that the cholera toxin is cased by the bacterium *Vibrio cholerae* and causes the rapid loss of fluid. According to the buffer system above, the loss of bicarbonate results in a shift of the bicarbonate buffer system to the side of the acid (H^+). If the volume of enteric fluid that is lost is large enough to overwhelm the ability of the kidney to regulate the proper acid-base levels in the body, **metabolic acidosis** results. An adult with severe untreated cholera in the blood pH can drop to a more acidic value of 7.19. To try to compensate, a clinical feature of metabolic acidosis is **hyperventilation** to try to eliminate the increased CO_2 more quickly, as by lowering the concentration of CO_2 in the body, the levels of H^+ are also lowered.

1.15 Male Reproduction

From an endocrinology standpoint, we have the following pathway:

Hypothalamus releases GnRH → Anterior Pituitary releases LH and FSH → Gonads produce steroids and germ cells (7)

Note that this general signaling pathway applies to both males and females. **GnRH** is the **gonadotropin releasing hormone**, **LH** is **luteinizing hormone**, and **FSH** is follicle stimulating hormone. Both LH and FSH are called **gonadotropins**. In males, the germ cells ultimately produced are called **spermatozoa**, while in the female they are the **ova**. The male reproductive anatomy is shown in Fig. 1.15. Because spermatogenesis requires a lower temperature than that of the body, they are produced in the testes that lie in the scrotum outside the body cavity. Within each testis are a series of convoluted tubules called the **seminiferous tubules**, and each of these tubules has the **spermatogenic cells**. The testes also produce the hormone **testosterone** using specialized interstitial cells called **Leydig cells** (also called the **interstitial cells**) that lie outside the seminiferous tubules. There are also **Sertoli cells** that lie outside the area of the seminiferous tubule, which promote spermatogenesis and produce the protein hormone **inhibin**. The Sertoli cells are in direct contact with the spermatogenic cells, and are adjacent to the luminal space.

Usually, the **spermatogonia** far away from the lumen of the seminiferous tubules will divide via mitosis. However, as

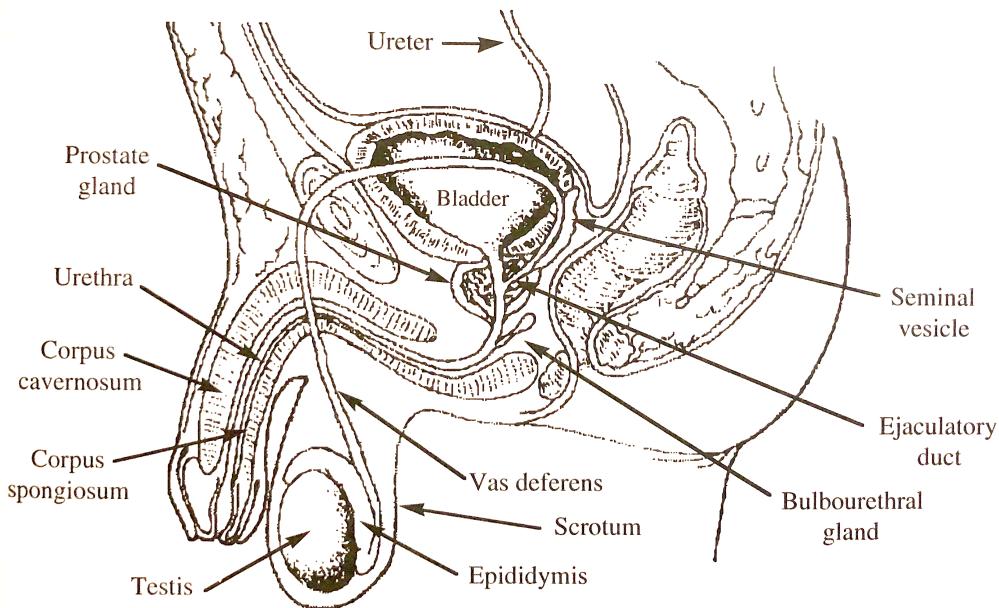


Figure 15: Male reproductive anatomy.

they get closer and closer to the lumen, eventually they are referred to as **primary spermatocytes** and begin to undergo meiosis, thus producing **secondary spermatocytes** after the first meiotic division and **spermatids** after the second meiotic division, which now have half as many chromosomes. The spermatids then transform into the **spermatozoa**, which are released into the lumen.

At the very tip of a spermatozoan, there is a structure called the **acrosome** which contains a lot of digestive enzymes that help the spermatozoa gain access to the interior of the egg once fertilization has taken place. Inside the head of the spermatozoa is a nucleus which contains the DNA. In the midsection are **mitochondria**, providing energy for the whipping movement of the tail to allow the sperm to swim towards their destination (i.e. the egg).

The Leydig cells produce testosterone by converting cholesterol to it. Testosterone can (1) diffuse out of the Leydig cells and move to other target tissues of the body, (2) diffuse into the nearby Sertoli cells, which convert testosterone into **dihydrotestosterone (dHT)**. dHT diffuses into the nucleus of the Sertoli cell and causes transcription that ultimately affects the spermatogenic cells.

LH (luteinizing hormone) binds to a specific receptor on the membrane of the Leydig cell, and causes the increase of the conversion of cholesterol into testosterone. Meanwhile, FSH (follicle stimulating hormone) binds to a surface receptor on the Sertoli cell that causes (1) an increase in the conversion of testosterone into dHT and (2) an increase in the synthesis of these FSH-sensitive receptors.

Testosterone has a *negative feedback* at the anterior pituitary and the hypothalamus, so it's important that it's able to diffuse out of the Leydig cells and enter the bloodstream. Specifically, it prevents the synthesis of GnRH and LH. Meanwhile, the Sertoli cells secrete a protein hormone called **inhibin**, which acts as a negative modulator of the anterior pituitary. Thus, if testosterone levels are too high, the feedback mechanism will decrease the levels of LH. If the levels of dHT are too high, there is an increase in inhibin synthesis that results in the decrease in FSH levels.

Almost all of the secondary sex characteristics in the male are due to testosterone. After sperm is produced, they leave the seminiferous tubules and enter into the **epididymis** and then the **vas deferens**. Sperm is stored in these regions for about 14 days before ejaculation. Contraction of smooth muscle lining the walls of these structures ejects the sperm down the vas deferens and into the **ejaculatory duct**. The **seminal vesicles**, **prostate gland**, and **bulbourethral gland** secrete components that make the remainder of the fluid that is ejaculated (now called **semen**). Some of the components secreted by these glands are fructose vitamins, bicarbonate zinc, prostaglandins, and mucus. Overall, **Fig. 1.15** shows us the general hormone involvement in the regulation of testes function.

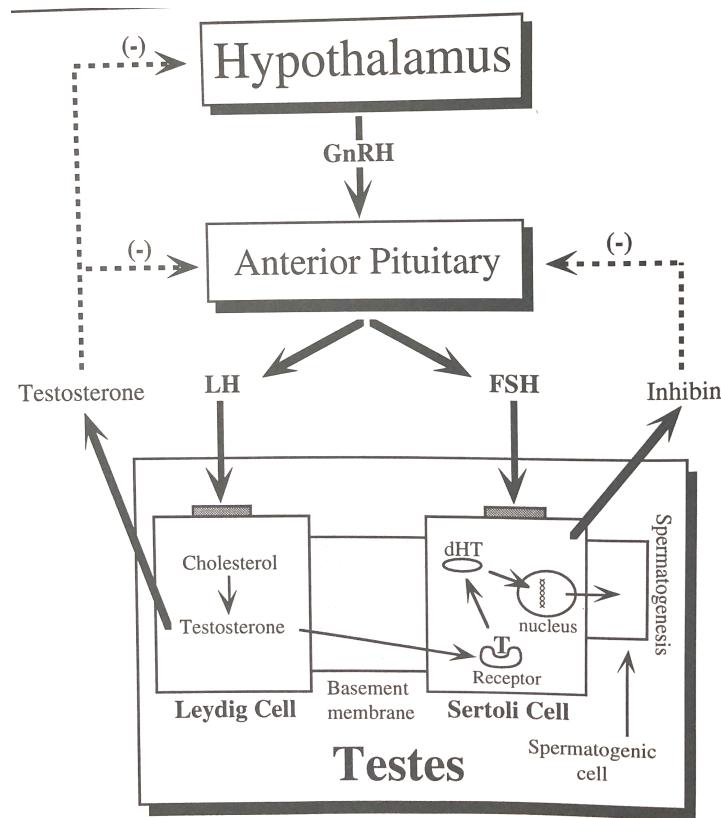


Figure 16: Hormonal regulation in the testes.

1.16 Female Reproduction

Once sperm have been ejaculated into the vagina, they can live for about 48 hours. About half an hour after ejaculation, the leading sperm arrive at the **Fallopian tubes** (also called the **oviduct**), which is typically the site of fertilization. The fertilized egg continues to move down the oviduct towards the uterus, where it will implant in the uterine lining. About a week after ovulation the fertilized egg, now called a **blastocyst**, implants in the lining of the uterus where it will continue to grow and develop until **parturition** (delivery). The general female reproductive anatomy is shown in Fig. 1.16. The production of female germ cells is called **oogenesis**. First, oogonia mitotically divide to form **primary oocytes**. All of these divisions to form primary oocytes occur in the first *two to three months of fetal development*. At birth a female will have about 400,000 primary oocytes, but by the time she has finished her reproductive life only about 400 will have ever matured. The remaining primary oocytes will have degenerated at various stages during the developmental process—the process of degeneration is called **atresia**.

The primary oocytes then undergo the first meiotic division, caused by a surge in the gonadotropin **LH**, forming the **secondary oocytes**. This first meiotic division happens in monthly cycles. During this division, one of the two daughter cells obtains all the cytoplasm while the other is just a small sphere of DNA called the **first polar body**. The secondary oocyte undergoes the second meiotic division *only after fertilization has taken place*. The result of that division is the **ovum** and a **second polar body**. The first polar body also divides to give two second polar bodies as well.

If we take a cross-section of the ovary, we will see a series of specific cell types called **primary follicles** that are in different developmental states. The primary follicle is a **primary oocyte** surrounded by a layer of follicle cells. These follicle cells are in constant contact with the primary oocyte. Eventually, one of the primary follicles will start to develop (two of them if there are fraternal twins). In order to mature the primary oocyte, **estrogen**, **LH**, and **FSH** are needed. This developing time period is referred to as the **follicular phase** and lasts up to about the 14th day of the woman's monthly cycle. During this period, the primary follicle gradually develops. Surrounding the primary oocyte will be a membrane called the **zona pellucida**, which itself is surrounded by more follicle cells called **granulosa cells** and finally by **theca cells**. These cell types are a direct result of the estrogen, LH, and FSH that are present. The **theca cells** are analogous to the **Leydig cells**

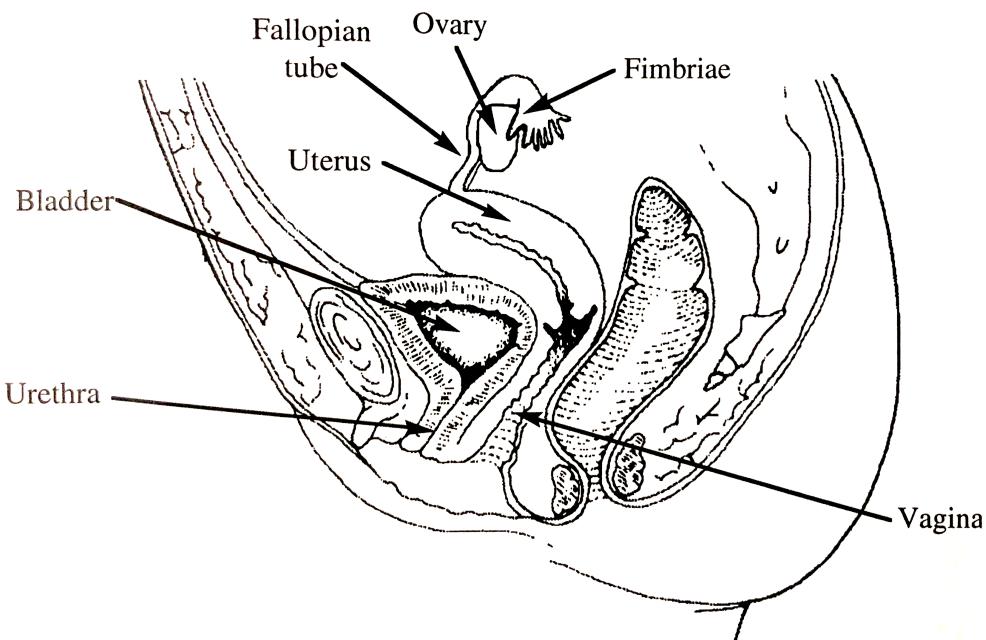


Figure 17: Female reproductive anatomy.

in the male, while the **granulosa cells** are analogous to the **Sertoli cells** in the male.

Within the primary follicle, a fluid starts to build up forming the **antrum**. Next, the **LH surge** event causes the primary oocyte to undergo the first meiotic division, forming a polar body and secondary oocyte. The LH surge also causes the production of a series of enzymes that break down the membrane in the primary follicle. Once the secondary oocyte is released, **ovulation** has occurred. Ovulation usually comes at about the 14th day in the woman's monthly cycle. The left over follicle after the completion of the follicular phase is transformed into a gland-like structure called the **corpus luteum**. One of the main functions of the corpus luteum is to produce **estrogen** and **progesterone**. If fertilization does not occur and there is no pregnancy, the corpus luteum will degenerate and the whole cycle starts again. From the point of ovulation, at about the 14th day, until the beginning of the menstrual flow is the **luteal phase**. See Fig. 1.16.

As we've mentioned previously, the **theeca cells** convert cholesterol into **testosterone**, which diffuses into the follicle cells where it is converted into **estrogen**. Estrogen helps in the development of the primary follicle (i.e. the primary oocyte). Meanwhile, recall that the anterior pituitary secretes LH and FSH—LH affects the theca cells while FSH affects the follicle cells. As estrogen is being synthesized, the primary oocyte is developing so we have a proliferation of follicle cells. If we have more follicle cells, then we can synthesize more estrogen. This is where the first estrogen surge (the one in the follicular phase) comes from. It turns out that at low concentrations of estrogen there is a negative feedback on FSH production. However, as the follicle cells are growing, they reach a certain threshold level of estrogen where at these high concentrations, estrogen actually has a *positive feedback* on the anterior pituitary and LH production. Thus, with a high level of estrogen production, we can now have the **LH surge**.

Immediately following the LH surge, the levels of LH drop down to very low levels. This is because at ovulation the follicle cells become transformed into the **corpus luteum**, meaning they are no longer follicle cells and hence transiently lose their ability to produce estrogen. Since we no longer have the high concentration of estrogen, we lose the positive feedback on the anterior pituitary.

Eventually, the corpus luteum becomes an endocrine gland and begins to synthesize estrogen and progesterone. Together, these hormones have a negative feedback on LH and FSH production in the anterior pituitary, and also on the hypothalamus on GnRH synthesis. Essentially, this prevents the primary follicle from developing, which is good because at this point there's no reason to have another primary follicle made. Remember, estrogen (in high concentrations) has a positive feedback, while estrogen and progesterone in the combination have a negative feedback.

During pregnancy, things are a bit different. The acrosome of the spermatozoan allows for the digestion of the membrane of the secondary oocyte. After the nucleus of the sperm enters the secondary oocyte, the **zona pellucida** changes and

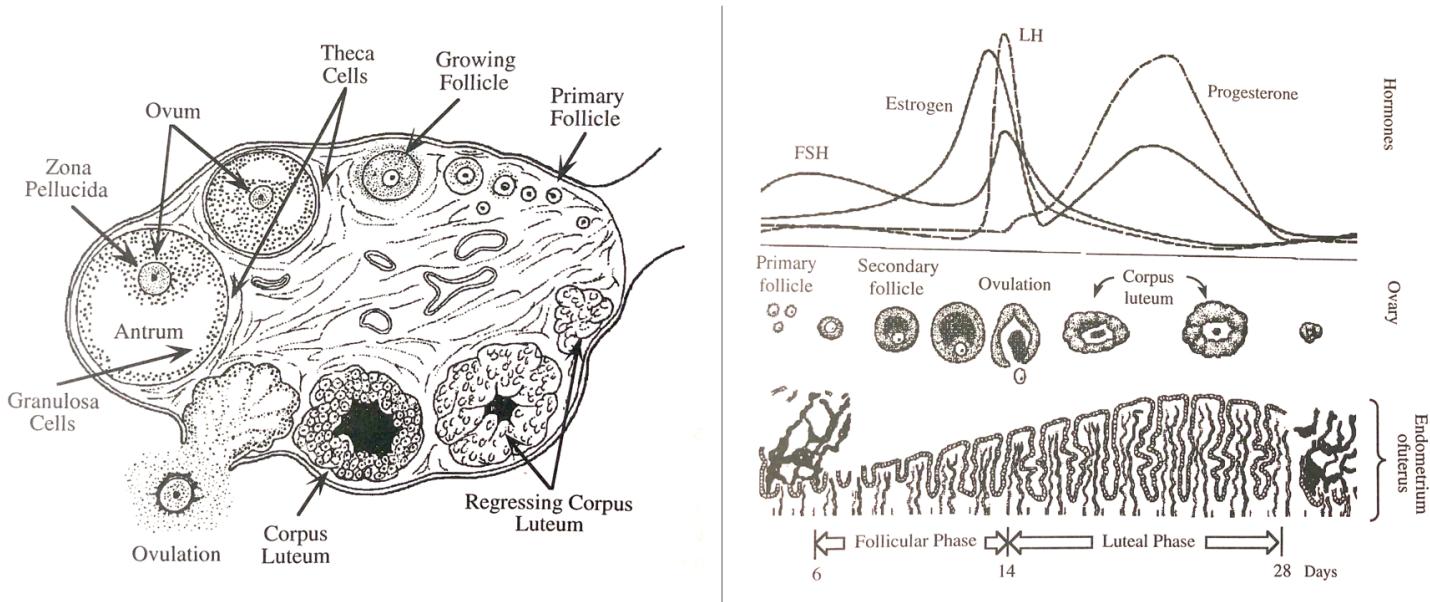


Figure 18: The ovarian cycle and its associated hormones.

prevents any other spermatozoa from entering—this process is referred to as **fertilization**. At this point the secondary oocyte undergoes the second meiotic division to form the **ovum** and the polar body. The nucleus of the sperm and egg fuse to form the **zygote**, which now has 46 chromosomes. The zygote rapidly develops and in about 7 days attaches itself as a **blastocyst** to the **uterine lining**. When the implantation takes place the **placenta**, made up of maternal and fetal cell types, begins to form.

For obvious reasons, we don't want the development of another primary follicle during pregnancy. Thus, throughout pregnancy, there are high levels of estrogen and progesterone. These high levels are maintained during the last six months of pregnancy, but for different reasons as we will discover.

1. During the first three months, the corpus luteum is still a viable gland and secretes estrogen and progesterone. The placenta is also an endocrine gland and synthesizes **chorionic gonadotropin (CG)**, which stimulates the corpus luteum to make estrogen and progesterone. CG is only made during the first three months of pregnancy. During this time, estrogen and progesterone work together to prevent the formation of primary follicles (and hence ovulation and menstruation as well) during pregnancy. The **mammary gland** is also sensitive to certain hormones during this time, such as **prolactin** that comes from the anterior pituitary and acts in a positive feedback manner on the mammary gland, and also **chorionic somatomammotropin (CS)** which comes from the placenta. CS and prolactin basically help the mammary gland grow. Estrogen and progesterone has positive feedback on the mammary glands as well.
2. During the last six months of pregnancy, CG is no longer made, so there is a loss of feedback to the corpus luteum. After three months, the corpus luteum breaks down and the supply of estrogen and progesterone from it comes to a halt. However, at this point in time the placenta itself starts to synthesize estrogen and progesterone.

1.17 Development

The **coelom** is the body cavity, basically the thoracic cavity and abdominal cavity. After fertilization, we have **cleavage** of the zygote, where the zygote rapidly divides into many smaller cells without an overall increase in size. **Gastrulation** then follows, where the cells of the zygote move to form the three primary germ layers (ectoderm, mesoderm, and endoderm) of the organism. The next developmental stage is **neurulation**, where we begin to see the formation of the nervous system, the first organ system to begin differentiation. **Neural crest** formation is the fifth developmental stage, which helps to form parts of the nervous system, skull, and sensory organs. The last stage of development is **organogenesis**.

Fertilization → Cleavage → Gastrulation → Neurulation → Neural Crest Formation → Organogenesis

(8)

Let's take a look at development in the frog to see an example of vertebrate development. In the unfertilized egg there is a large amount of yolk that resides in the **vegetal pole** (i.e. the lower hemisphere), which will act as food for the developing embryo. The **animal pole** (i.e. the upper hemisphere) contains mainly cytoplasm. The **gray crescent** is located on the **dorsal aspect** (i.e. the future back) of the egg, opposite to where the sperm penetrates the egg's membrane. It eventually is where we will find the spinal cord and brain.

Once the zygote has formed, it undergoes special cell division to increase its mass but not its overall size. The individual cells involved in this growth are called **blastomeres**. Eventually, a small, solid ball of cells will be formed called a **morula**. Further cellular division results in the formation of a hollow ball of cells called the **blastula**, which includes a fluid-filled cavity called the **blastocoel**. These different initial stages are shown in Fig. 1.17.

During **gastrulation**, rearrangement of cells occurs. Not far from the gray crescent, an opening develops in the blastula

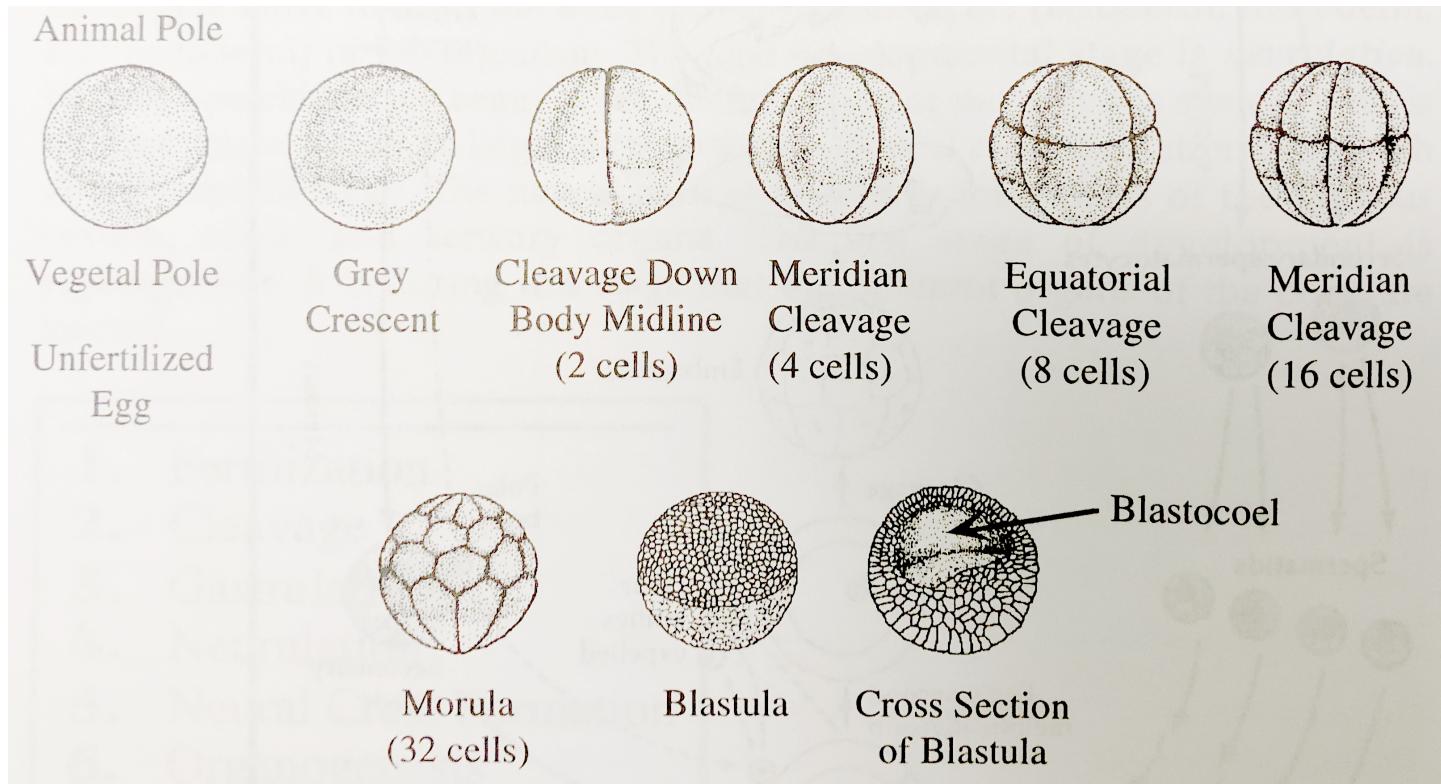


Figure 19: Different stages in embryonic development.

called the **blastopore**. Cells from the animal pole begin to migrate inwards through the **dorsal lip** of the blastopore. As this outer layer of cells migrates inward, they form a second layer of cells immediately below. The blastocoel is reduced in size and is ultimately eliminated—in its place, a new cavity is formed called the **archenteron**. At this stage, the embryo is referred to as a **gastrula**. Invagination of the outer cell layer produce two cell layers: the **ectoderm** (outer layer) and **endoderm** (inner layer). A layer of **mesoderm** will later form between these two cell layers. The different stages of gastrulation are shown in Fig. 1.17. In general, it seems that the generation of three cell layers has been conserved through the evolutionary process.

Here are important *developmental fates* to remember:

1. **Endoderm:** inner lining of the digestive and respiratory tracts, and major glands of the body like the liver and pancreas.
2. **Mesoderm:** notochord, heart, skeleton, muscle, outer coverings of internal organs, and reproductive organs.
3. **Ectoderm:** skin, lens of the eye, brain and nervous system.

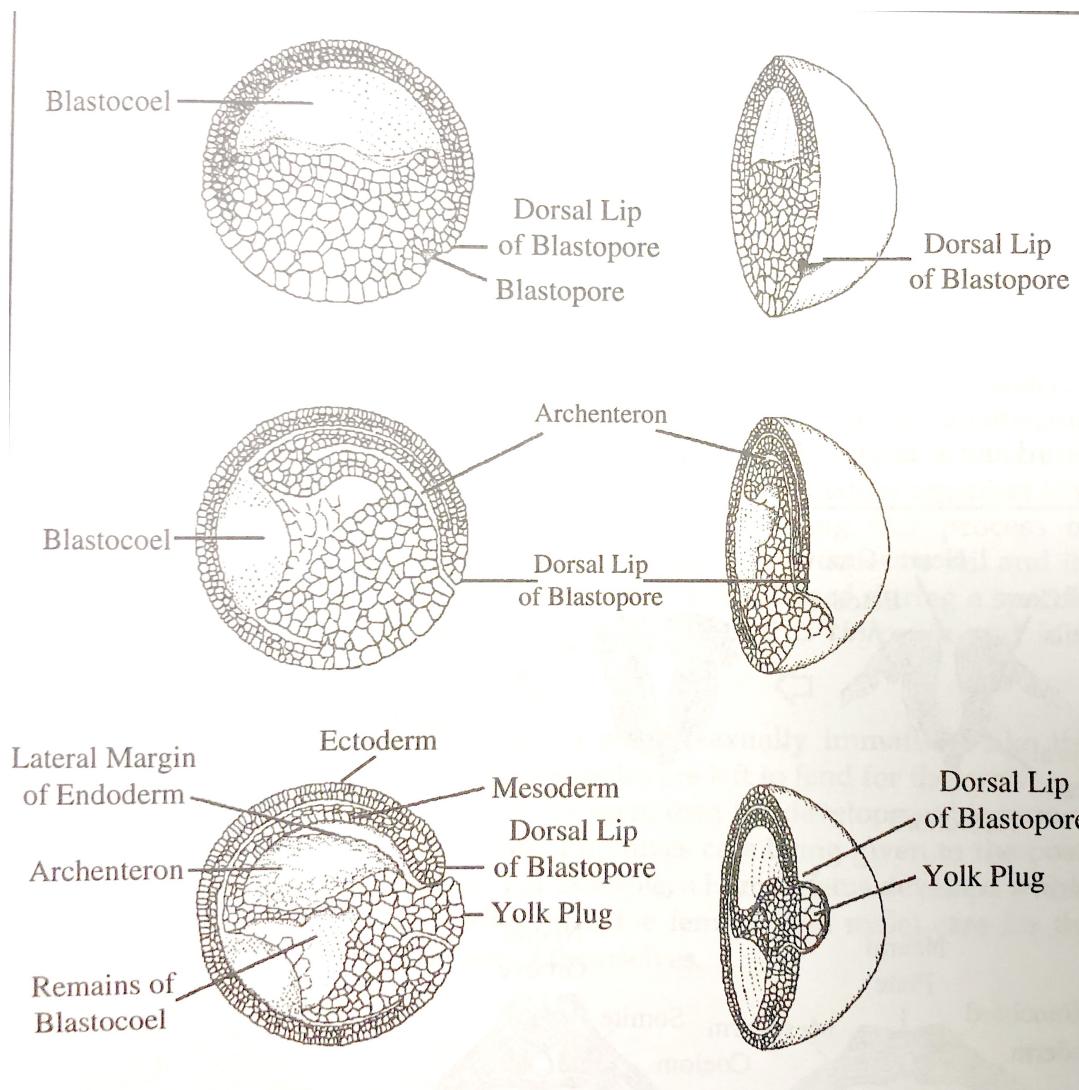


Figure 20: The process of gastrulation.

These are just a few of the developmental fates of these cell layers. Now, during the next stage of development, **neurulation**, the ectoderm, mesoderm, and endoderm begin to form the structures that will eventually define the embryo and later the adult. Notably, the formation of the **notochord** takes place along the body midline, and is derived from the *mesoderm*. Superior to the notochord is a mass of *ectodermal cells* called the **neural plate**, which begins to fold in on itself and form the **neural groove**. As the neural groove forms, the mesoderm is also split and forms the coelom (i.e. the body cavity). The neural groove goes on to form the **neural tube**, which will eventually encase the spinal cord (encased in the *spinal column*), and anterior to the spinal cord will form the brain. In other words, the neural plate, which is composed of ectodermal cells, gives rise to the nervous system. This tissue is referred to as **primordium**, and the embryo is referred to as a **neurula** during this stage of development. **Fig. 1.17** shows the formation of the neural tube during neurulation.

When the edges of the neural groove fused together and become the neural tube, specialized ectodermal cells were left outside the tube and are called **neural crest cells**. These cells begin to form various parts of the body, such as the sensory cells of the head, the adrenal medulla, and other things. This process is called **neural crest formation**.

Organogenesis first begins with interactions between the ectoderm and mesoderm. The neural tube becomes longer and thinner. Mesodermal cells migrate towards the neural tube and eventually form the vertebral column. The brain and optic vesicles begin to form. As the neural tube continues to form, segments of mesodermal tissue called **somites** begin to appear, which eventually gives rise to the vertebrae, connective tissue, and muscles of the body.

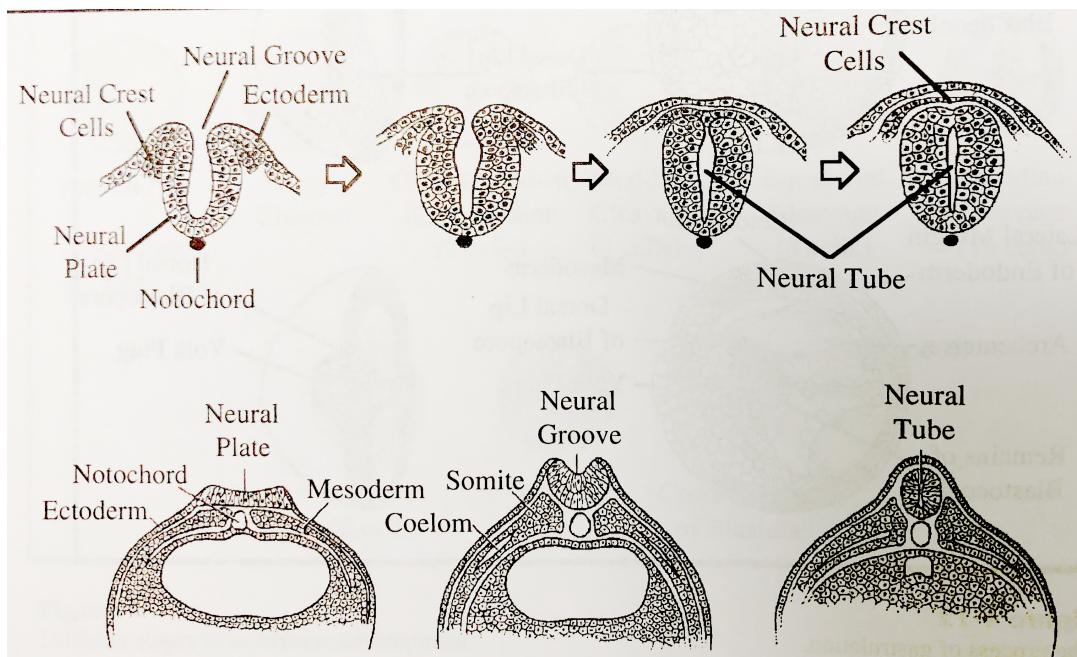


Figure 21: Neural tube formation.

1.18 Developmental Mechanisms

There are two general classes of interaction associated with the differentiation of cells that we need to consider: intracellular and intercellular interactions. Intracellular interactions usually result in the setting up of a *prepattern*, while the intercellular interactions usually undergo *developmental induction*.

Recall that after fertilization and induction of development, the gray crescent forms at a point opposite to where the sperm penetrated the egg. The formation of the gray crescent is due to an intracellular interaction, and is not a prepatterned phenomenon simply because of the fact that the entry of the sperm in the animal half of the cell is random. However, once entry is made and the gray crescent forms, the dorsal midline is established. This allows us to define directions. In other words, prior to the first cellular division, the axes of the organism are already established. Hans Spemann showed that the gray crescent is essential for complete embryonic development in the 1920s.

1.19 Human Embryo Development

Recall that the **blastocyst** is a hollow ball of cells with a mass of cells on one side. The surrounding cells that form the ball are called the **trophoblast**, and the other are called the **inner cell mass**. If the lining of the uterus is prepared to receive the embryo, then it will implant. Otherwise, the embryo is rejected and sloughed off during the next menstrual period. After implantation, the trophoblastic cells grow into the lining of the uterus. (In fact, the cells of the trophoblast grow little finger-like projections into the uterus.)

For nourishment, blood vessels from the umbilical cord grow into the projections of the trophoblast. There is a layer of cells between the trophoblast and the fetal blood vessels called the **chorion**, which preserves the barrier between the mother's and fetus's blood (but diffusion of nutrients and waste products may still occur). This whole exchange apparatus is called the **placenta**.

The **inner mast cells** will undergo changes similar to those in the frog to form the three basic germ layers. We also have the **primitive streak**, which is equivalent to the neural plate in frogs, that forms in the ectoderm. Cells from the primitive streak migrate down between the ectoderm and endoderm to become the mesoderm. Further folding of the primitive streak gives rise to a neural groove and then a neural tube, etc. So, the formation of the primitive streak in mammals marks the beginning of gastrulation, which is quickly followed by neurulation. Cell fate determination from the three germ layers in

humans is basically the same as in frogs.

The embryo cannot implant if the uterus is not receptive—that is, if it is not **quiescent**. Furthermore, pregnancy will not be maintained by the uterus even after implantation unless it remains quiescent. So, to maintain the uterine lining through the first part of pregnancy, the trophoblastic cells secrete **chorionic gonadotropin (CG)**, which is a hormone that causes the corpus luteum to continue to produce estrogen and progesterone. Eventually, the placenta can take over the production of estrogen and progesterone itself and the corpus luteum is no longer needed.

The placenta secretes increasing amounts of estrogen and progesterone, but the levels of estrogen increase faster than those of progesterone. At some point near the end of pregnancy, the levels of progesterone plateau, and at a critical ratio of [estrogen]:[progesterone], the uterus is no longer quiescent; it begins to have contractions. The smooth muscle of the uterus contracts, putting the walls of the uterus under tension that thus sends nerve impulses to the hypothalamus. The hypothalamus then sends signals to the posterior pituitary to release **oxytocin**, which is a strong inducer of more contractions of the smooth muscle of the uterus. Oxytocin also stimulates the production and secretion of prostaglandins that further induce contractions. All of these hormones and nerves form a *positive feedback system*. We call the birth process **parturition**.

Milk production and milk ejection are two important processes in lactation. There are epithelial cells within the breast that make up the glands that produce the milk, and around these cells, we have the myoepithelial cells that can contract around the milk glands, ejecting the milk. Production of milk is stimulated by the suckling of the infant on the mother's nipple. The nipple sends a nervous impulse to the hypothalamus, which responds by releasing **PRH (prolactin releasing hormone)**, which acts on the anterior pituitary. PRH stimulates the anterior pituitary to release **prolactin**, which goes into the bloodstream and stimulates the epithelial cells in the breasts which comprise the milk glands to produce more milk. Suckling on the nipple also tells the hypothalamus to send a nervous signal to the posterior pituitary and release **oxytocin**, which causes contraction of the myoepithelial cells in the breast to squeeze the milk out.

1.20 Endocrinology

Hormones can generally be divided into *peptide* hormones (e.g. insulin), *amine* hormones (e.g. epinephrine or adrenaline, which are classified as **catecholamines**), and *steroids* (e.g. progesterone and estrogen). *Catecholamines* are monoamine neurotransmitters, which are organic compounds that have a catechol (benzene with two hydroxyl side groups next to each other) and a side-chain amine. Hormones are released by endocrine organs into the blood and travel by way of the circulatory system to various target tissues. To illustrate the importance of hormones, let's consider the action of the catecholamine epinephrine (adrenaline) on a typical hepatic (liver) cell. When the body is under some type of stress like physical exercise or even fright, an increased need for glucose arises.

Once a stress has been perceived, the nervous system responds by signaling the adrenal medulla (part of the adrenal gland that sits on top of the kidneys) to release epinephrine into the extracellular fluid. Epinephrine diffuses into the blood, and the **β -adrenergic receptors** on hepatic cells bind to epinephrine and cause the activation of **adenylyl cyclase** (bound on the cytoplasmic membrane surface), which increases the concentration of cAMP (cyclic adenosine monophosphate, a second messenger) in the cell. β -adrenergic receptors are a type of *G-protein coupled receptors*. GTP-bound state corresponds to the active state, while GDP-bound state corresponds to the inactive state. The increase in cAMP concentration allows cAMP to interact with *protein kinase A (PKA)* and help it phosphorylate an enzyme called **glycogen phosphorylase**. Once glycogen phosphorylase has become phosphorylated, it is now an active enzyme and catalyzes the conversion of glycogen into glucose.

Cholera is an intestinal disorder caused by the *Vibrio cholerae* bacterium. The major symptom of this disorder is **diarrhea**, and if left untreated will result in severe dehydration and eventual death. This toxin binds to the active state of the G protein and prevents GTP from being hydrolyzed to GDP. This means that the adenylyl cyclase enzyme is continually active and massive amounts of cAMP are synthesized. cAMP causes the intestinal cells to secrete digestive fluids.

Let's consider another example, this time involving the water-soluble peptide hormone **gastrin** that stimulates the secretion of HCl and pepsinogen from the stomach in response to stimulation from the vagus nerve and partially digested protein. Gastrin first binds to a GPCR, activating an associated G-protein that can now interact with the membrane enzyme **phospholipase C (PLC)**, and this interaction induces PLC to hydrolyze phosphatidyl-inositol-4,5-biphosphate (PIP₂) to inositol-1,4,5-triphosphate (IP₃) and 1,2-diacylglycerol (DAG). IP₃ is a second messenger that interacts with the IP₃-sensitive calcium ion channels in the endoplasmic reticulum membrane, stimulating the release of Ca²⁺ ions into the cytosol from the ER lumen. Meanwhile, DAG, diffusing through the plasma membrane, interacts with **protein kinase C** and stimulates that kinase with the help of Ca²⁺ to phosphorylate an unknown protein which in turn causes HCl secretion into the lumen of the stomach.

Another example of a peptide hormone in action involves insulin, which is a water-soluble peptide hormone that binds to a specific trans-membrane receptor in the cell membranes of liver, fat, and muscle cells. Once insulin binds to the receptor

on the cell surface, the cytoplasmic portion of the receptor is converted into a tyrosine kinase that autophosphorylates the amino acid tyrosine found within the cytoplasmic portion of the receptor. This acts to further enhance the activity of the tyrosine kinase. Presumably, the insulin receptor can also internalize and somehow act as a second messenger. This action, along with enhanced tyrosine kinase activity, leads to the internalization of glucose into the cells. The actual events in the insulin signaling mechanism that leads to the uptake of glucose are somewhat obscure at the present time.

While these show a lot about some of the many different pathways in cell signaling, kind of the main thing to pay attention to is that peptide and catecholamine hormones always involve some signaling intermediates because they can't diffuse freely through the membrane. Contrast this with steroid and thyroid hormones, which are lipid soluble hormones that can pass through the cell's plasma membrane and interact with a receptor either in the cytosol or in the nucleus. **Thyroid hormones** can diffuse across the plasma membrane and into the nucleus where they bind with specific receptors. The hormone receptor complex then activates transcription essential for certain metabolic processes; indeed, thyroid hormones help to regulate growth and differentiation and they can stimulate the breakdown of proteins, fats, and glucose.

There are four major types of regulatory mechanisms in endocrinology: endocrine, neuroendocrine, paracrine, and autocrine. We go over an example of each here:

- **Endocrine Regulation:** The **pancreas** secretes insulin and glucagon, which are two hormones that are important in maintaining the proper levels of blood glucose. Both of these hormones are secreted from clusters of specialized cells called the **islets of Langerhans**. Insulin is secreted by the β cells, while glucagon is secreted by the α cells. An increase in blood glucose stimulates the β cells to produce insulin and excrete it into the blood, which goes to the liver, fat, and muscle cells and tells them to uptake glucose, thus decreasing blood glucose levels. Similarly, a decrease in blood glucose stimulates the α cells to produce glucagon and excrete it into the blood, which goes to the lever and fat cells and tells them to release glucose and fatty acids, thus increasing blood glucose levels. This is an example of **negative feedback**.
- **Neuroendocrine Regulation:** In this case, the hormone is not released for an endocrine cell, but rather from a nerve cell which releases its neurotransmitter in the form of a hormone into the blood. For example, the **adrenal medulla** can receive sensory input from a sympathetic nerve, which tells it to release epinephrine into the blood. Other examples of neuroendocrine regulation involve the **hypothalamus** and the **pituitary gland**. The pituitary gland can be divided into the **anterior** and **posterior pituitary**.
- **Paracrine Regulation:** In paracrine regulation, the chemical that acts as a signal is released from one cell and influences a cell immediately adjacent to it. An example of such a paracrine cell would be **mast cells**, which contain large amounts of **histamine**. The substances released by the paracrine cell are generally dumped into the extracellular space and not into the bloodstream. Other examples of a paracrine signal would be neurohormones and neurotransmitters.
- **Autocrine Regulation:** In autocrine regulation, cells can release certain chemicals which they can then respond to themselves. For example, certain cells can release growth factors which can then bind to specific receptors on the membrane of that same cell. Thus, the cells that released the growth hormone are stimulated to grow.

1.21 The Pituitary Gland

The posterior pituitary releases **oxytocin** and **antidiuretic hormone**, which are synthesized in specific cells of the hypothalamus. Oxytocin stimulates female uterine contraction in a positive feedback loop while antidiuretic hormone (ADH) stimulates water and Na^{2+} reabsorption in the kidneys and also helps to increase the blood volume/pressure. The ADH that is synthesized in the nerve cell bodies in the hypothalamus are packaged into vesicles and transported down the axon to the terminal bouton in the posterior pituitary. A nerve impulse propagates down that same axon and causes the release of these hormones into a system of nearby capillaries. ADH helps prevent **diuresis**, which is the excessive loss of urine.

The anterior pituitary secretes six major hormones: thyroid stimulating hormone (TSH), adrenocorticotrophic hormone (ACTH), follicle stimulating hormone (FSH), luteinizing hormone (LH), growth hormone (GH), and prolactin (PRL). These hormones are regulated by a second set of hormones stored in the hypothalamic nerves. For example, the hypothalamus has nerve cells that contain thyrotropin releasing hormone (TRH). When this nerve is stimulated, it secretes TRH into a set of capillaries which extend into the anterior pituitary. TRH stimulates the synthesis and release of TSH, which in turn binds to specialized receptors in the thyroid gland and causes the release of thyroxine. Thyroxine increases the rate of metabolism and growth as a part of a negative feedback loop.

1.22 Immunology

Blood contains **erythrocytes** (red blood cells), **leukocytes** (white blood cells), and platelets. Erythrocytes are produced in the marrow of the sternum, ribs, and vertebrae, while leukocytes are produced partially in the tissues of the lymph and partly in the bone marrow.

There are six types of leukocytes found in the blood, but we are only interested in three of them: **monocytes**, **neutrophils**, and **lymphocytes**. Monocytes and neutrophils are considered to be **phagocytes**, which, along with **mast cells** and a variety of other cell types, participate in the *immune response*.

- **Mast cells** are derived from leukocytes and then migrate out into the tissues where they reside. When mast cells are stimulated, they release **histamine** which acts on endothelial cells and causes an increased permeability to cells like neutrophils, allowing neutrophils easy access to the surrounding tissue in order to defend against foreign pathogens.
- **Phagocytes** include monocytes (i.e. **macrophages** when they leave the blood and enter into the tissues) and neutrophils. These cells are the primary cell types that attack and destroy foreign bacteria and viruses. They do this by the process of **phagocytosis**, engulfing the foreign invader by **endocytosis**.
- Lymphocytes are derived from either the **thymus** gland (which produces **T lymphocytes**) or elsewhere from places that produce **B lymphocytes**. T cells are responsible for *cell-mediated immunity*. These cells are responsible for the destruction of foreign microorganisms and other harmful agents. There are three types of T cells: cytotoxic/killer T cells, helper T cells, and suppressor T cells. B cells are responsible for *humoral mediated immunity*. Upon infection, B cells can differentiate into **plasma cells** that can synthesize and secrete antibodies. **Antibodies** are proteins that are synthesized in response to an **antigen**, which is simply a foreign substance with high MW.

Let's go over the two types of general immune responses—cell-mediated immunity and humoral mediated immunity—in more detail:

- **Cell-Mediated Immunity:** Macrophages engulfs a foreign particle and phagocytizes it. The antigenic fragments released into the cytosol of the macrophage are transported to the macrophage's membrane, where they bind to the **major histocompatibility complex protein, class 1 (MHC I)**. Certain receptors on **cytotoxic T cells** recognize the antigen-MHC I complex on the macrophage and bind to it, stimulating the release of growth factor **interleukin-1 (IL-1)** from the macrophage. The cytotoxic T cell also releases **interleukin-2 (IL-2)**. IL-1, IL-2, and other interleukins released by **helper T cells** (discussed below) stimulate the synthesis of more cytotoxic T cells. These killer T cells proliferate and bind to the invading foreign cells bearing the antigen, causing them to lyse.
- **Humoral Mediated Immunity:** B cells have **MHC class 2 (MHC II) proteins** and antibodies on their surface. When a B cell finds an antigen that has specifically bound to its antibody, it engulfs that antigen-antibody complex and transports a portion of that antigen to the MHC II protein, effectively ‘displaying’ the complex on the surface of its membrane. **Helper T cells** with the right receptors are able to bind to the antigen-MHC II complexes, stimulating the helper T cells to release interleukins (a **lymphokine**). This stimulates the B cells to proliferate and form **plasma cells**, which in turn produce a vast amount of **antibodies** specific to the original antigen. When these circulating antibodies bind to the antigen, they act as a tag that signals circulating phagocytes to engulf the antigen-antibody complex and destroy it. The human immunodeficiency virus (HIV) acts at the level of the helper T cells by infecting them.

1.23 Antibodies

Most immunoglobulins (antibodies) are composed of 4 subunits arranged in a ‘Y’ configuration. There are 2 light chains and 2 heavy chains. These subunits are joined to one another by disulfide bonds. Within each heavy and light chain are **variable domains** and **constant domains**.

At the terminal ends of the heavy and light chains are variable (V) regions that can differ in amino acid sequence from immunoglobulin to immunoglobulin. The constant (C) regions of the heavy and light chains are found in the lower portions of the immunoglobulin. The **antigen binding site** for a particular antibody is located at the end of the variable regions of the heavy and light chains.

Two other regions of the immunoglobulin are important: the diversity (D) and the joining (J) region. We show a picture of the general antibody structure in **Figure 1.23**. There are five classes of immunoglobulins that differ in the composition of their heavy chains:

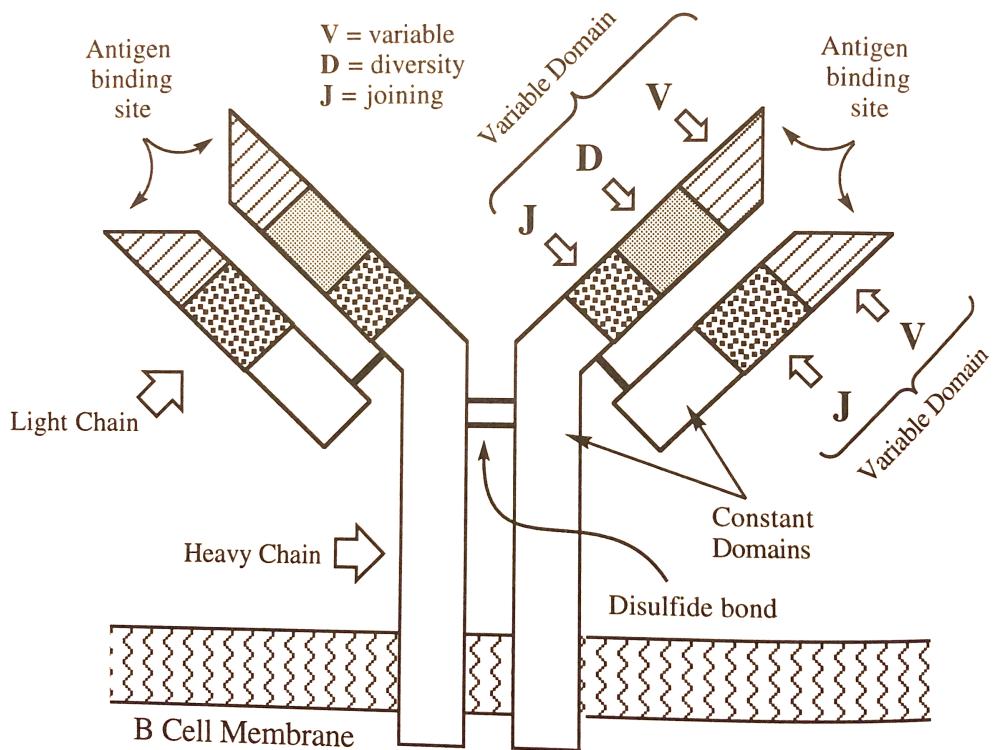


Figure 22: Generalized antibody structure.

- **IgA:** Found in milk and helps to protect nursing infants.
- **IgD:** Has unknown function.
- **IgE:** Binds to mast cells and is involved in the allergic reaction.
- **IgG:** The only antibody able to cross the placenta. It is also the most abundant and is produced within days after the IgM antibody is secreted.
- **IgM:** Produced a few days after detection of an antigen and it is the first antibody produced in response to an antigen.

Antibodies simply recognize and identify foreign antigens, and can have different mechanisms by which to accomplish this function:

- **Direct Blocking:** The antibody directly blocks the foreign invader from gaining access to host cells. This is accomplished by the antibodies binding to the antigens or other parts of the pathogen.
- **Complement:** An antibody has already recognized and bound to a specific antigen on a bacterial cell that is considered an intruder, and a complement protein (plasma protein) binds to the Fc domain of the antibody. After a series of reactions, the complement protein is activated and triggers an immune response. Further reactions form a **membrane attack complex (MAC)** that inserts into the bacterial cell's membrane and forms a channel that lets water into the cell, causing it to swell with water and eventually lyse.
- **Coating the Cell Surface:** Antibodies can bind to specific antigens on the surface of a bacterial cell and coat the cell surface. Once the antibodies have attached to the bacterial cell, phagocytes and/or killer T cells can bind to the terminal portion of the Fc domain of the antibodies and begin to engulf the foreign invader.

1.24 T Cells

The receptors for T cells are composed of two polypeptide chains, each with a constant and variable domain. Within a variable domain in each polypeptide we find a variable (V) region and a joining (J) region. On one of the polypeptide chains is a diversity (D) region. If a virus infects a host cell, then that virus will begin to take over the host's metabolic machinery.

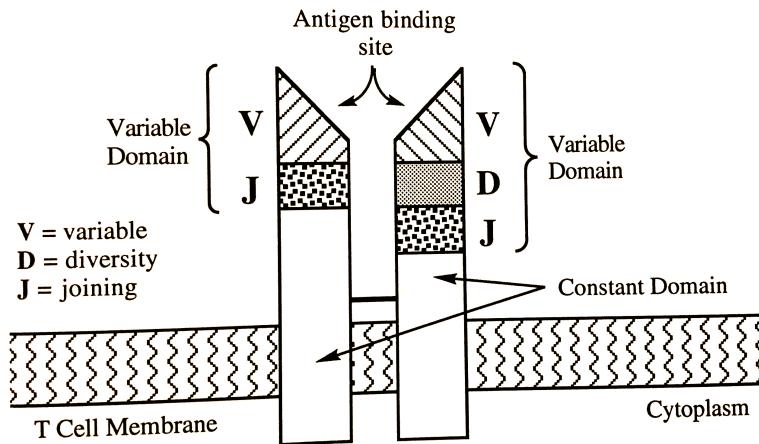


Figure 23: Example of T-cell Receptor Structure

As this happens some of the viral antigens are transported to the surface of the host cell where they can complex with a MHCI protein receptor, which are found on almost every one of our cells. The antigen-MHCI complex is recognized by the cytotoxic T cells, which induce lysis in the host cell. Immature helper T cells recognize macrophages which have presented an antigen on their MHCII receptors. Binding induces the macrophage to synthesize and release IL-1, which acts on the immature helper T cell and causes it to release IL-2, further stimulating the immature helper T cell to proliferate into a mature helper T cell. The mature form of the helper T cell secretes IL-2 which can activate cytotoxic T cells, B cells, and more helper T cells.

1.25 Immunology—Generalized Review

Let's summarize the basic events in humoral and cellular immunity:

1. A virus/bacterium enters the body by the blood and is engulfed by a macrophage.
2. On the surface of the macrophage are MHCI receptors and MHCII receptors.
3. A MHCII receptor presents the viral antigen to the receptor of a helper T cell, causing the macrophage to release IL-1 that stimulates the helper T cell to proliferate.
4. The helper T cell is stimulated to release IL-2 which enhances proliferation of helper T cells.
5. A B cell with a MHCII protein presents viral antigen to helper T cells, and the IL-2 released from helper T cells stimulates B cells to proliferate.
6. B cells produce plasma cells and memory B cells. Memory B cells 'remember' antigen and proliferate faster during a future invasion of the same virus. Plasma cells secrete antibodies specific for the viral antigen.
7. The antibodies respond by direct block, complement, and cell surface coating.
8. IL-2 from the helper T cells stimulates cytotoxic T cells which have bound to the MHCI protein-antigen complex of an infected cell to lyse the infected cell.
9. Interferon is secreted by the infected host cell and acts on the cytotoxic T cell to help enhance the immune response.
10. Cytotoxic T cell also make memory T cells, which will proliferate faster during a future invasion of the same virus.

Fig. 1.25 shows a diagram that basically summarizes the above information.

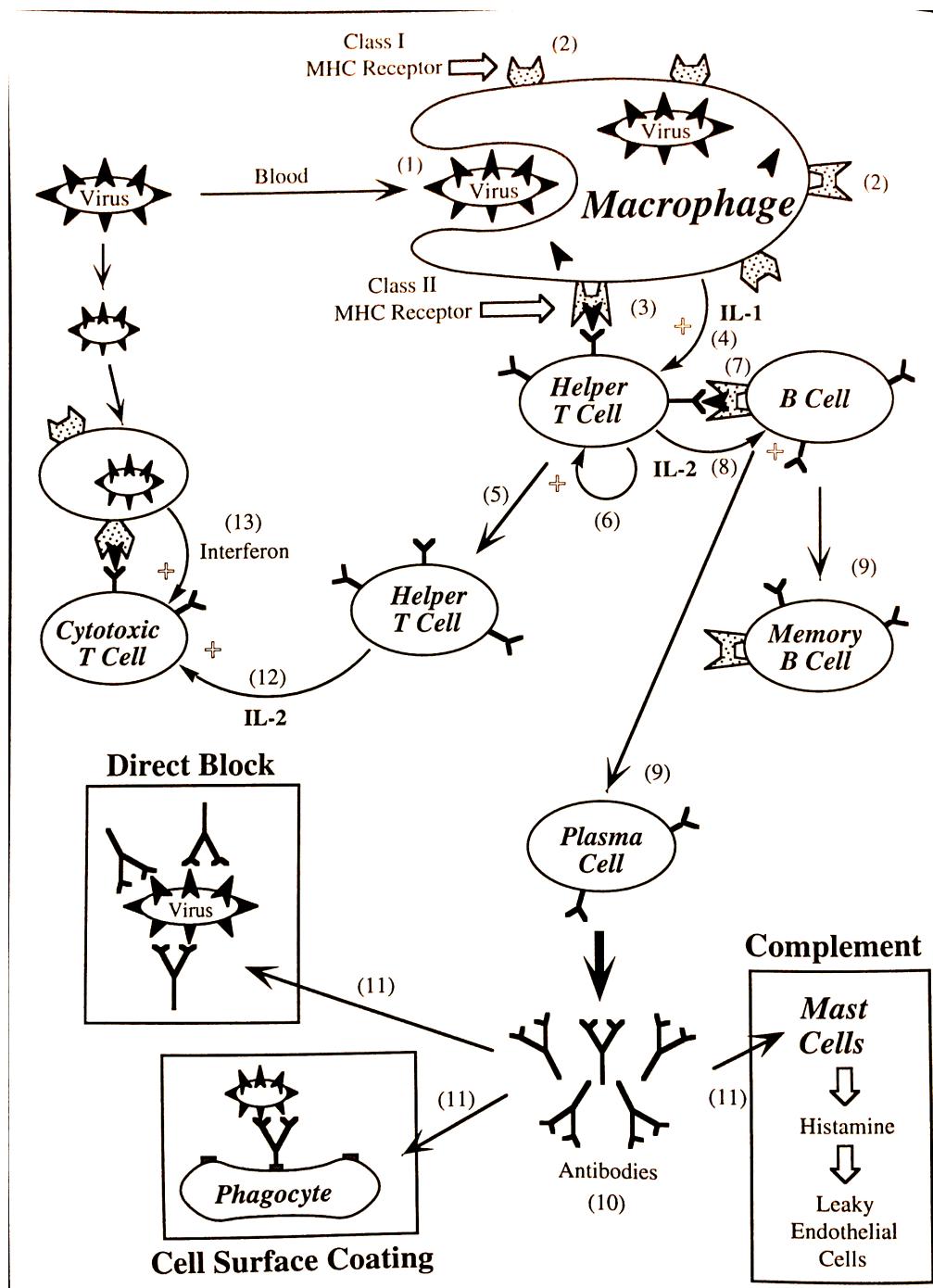


Figure 24: Review of humoral and cellular immunity pathways.

2 General Chemistry

2.1 Stoichiometry

The MCAT does involve some math, but often it isn't too complicated. Math-related calculations required for MCAT questions involve making approximations, determining ratios, setting up calculations, and estimating the effect of errors on results. Unit conversions to memorize:

- Distances: $1\text{ m} = 1.094\text{ yd}$, $2.54\text{ cm} = 1\text{ in}$, $1.609\text{ km} = 1\text{ mi}$
- Masses: $1\text{ kg} = 2.205\text{ lb}$, $453.6\text{ g} = 1\text{ lb}$
- Volumes: $3.79\text{ L} = 1\text{ gal}$, $1\text{ L} = 1.06\text{ qt}$
- Temperatures: $T_F = \frac{9}{5}T_C + 32$, $T_C = \frac{5}{9}(T_F - 32)$

Molality (m) is the concentration of a fluid solution defined as the moles of solute per kilogram of solvent. The molality of a solution does not change with temperature, so it is used to calculate the boiling-point elevation and freezing-point depression of solutions containing non-volatile impurities. **Mass percent** is the concentration of a fluid solution defined as the mass of the solute per mass of solution.

Beer's Law is defined to be $\text{Absorbance} = \varepsilon Cl$, where ε is a constant for the solute at λ_{\max} (the wavelength of greatest absorbance), C is the solute concentration, and l is the width of the cuvette.

Here are some important solubility rules to remember:

1. Most salts containing alkali metal cations (Li^+ , Na^+ , K^+ , Cs^+ , Rb^+) and ammonium (NH_4^+) are water-soluble.
2. Most nitrate (NO_3^-) salts are water-soluble.
3. Most salts containing halide anions (Cl^- , Br^- , I^-) are water-soluble (with heavy metal exceptions such as Ag^+ and Pb^{2+}).
4. Most salts containing sulfate anions (SO_4^{2-}) are water-soluble (with exceptions such as Ba^{2+} , Pb^{2+} , Hg^{2+} , and Ca^{2+}).
5. Most hydroxide anion (OH^-) salts are only *slightly* water-soluble. KOH and NaOH are substantially soluble, while $\text{Ca}(\text{OH})_2$, $\text{Sr}(\text{OH})_2$, and $\text{Ba}(\text{OH})_2$ are fairly soluble in water.
6. Most carbonate anion (CO_3^{2-}), chromate anion (CrO_4^{2-}), phosphate anion (PO_4^{3-}), and sulfid anion (S^{2-}) salts are only slightly water-soluble.

There's also some general test-taking tips and general advice that could prove very worthwhile:

1. Keep in mind that on a multiple-choice exam, the math has been done for you, so all you need to do is approximate the answer.
2. Intuition will prove useful on the MCAT. As you study for your MCAT, you need to learn to trust your intuition.
3. Keep in mind that you are *not* graded for showing your work on the MCAT, so don't solve every problem to the last decimal place. Analyze each question only well enough to eliminate three wrong answers. Be concise and efficient in your problem solving, not exhaustive.

2.2 Atomic Theory

There are a couple of classical experiments/lab techniques that helped to develop the modern atomic theory that you should be familiar with:

1. **Thomson Experiment:** This experiment helped determine the sign of charges and the charge-to-mass ratio of the electron. Basically, Thomson had a beam of electrons, and showed that he could deflect the beam by applying an electric field by an exact amount each time. The magnitude of deflection depended on the strength of the field, and reversing the plates of the external field got the same magnitude of deflection, but in the opposite direction. Thomson also figured out the charge-to-mass ratio of electrons to be $1.76 \times 10^8\text{ C/g}$, as the arc of deflection was constant in magnitude when the field strength was held constant but its orientation was aligned differently.
2. **Millikan Oil Drop Experiment:** This experiment helped determine the magnitude of charge of subatomic particles. The experimental setup involved trying to suspend a charged oil drop in an electric field. The oil drop became charged by adding an electron to it. The droplet is able to be suspended by balancing the force of gravity with the force produced by the electric field on the charged oil droplet. Mathematically, this looks like $qE + mg = 0$, so $q = -\frac{mg}{E}$. This helped determine the charge on an electron.

- 3. Rutherford Experiment:** This experiment helped determine the localization of mass/dense particles in an atom. A thin strip of gold foil lies perpendicular to an incoming stream of α -particles. The beam strikes the gold strip surrounded by a zinc sulfide ZnS band that luminesces if struck by α -particles. Rutherford observed that very few α -particles ricocheted off the gold, and most passed straight through the gold. This showed that most of an atom was empty space, and that atoms have dense nuclei with nearly all of the atomic mass centrally concentrated.
- 4. Mass Spectrometry:** A mass spectrometer is designed to measure the charge-to-mass ratio for a charged particle. This is accomplished by sending a particle into a perpendicular magnetic field and observing the degree to which it curves. By comparing the curvature for an atomic or molecular ion to a known standard, the mass of the unknown ion and/or the mass-to-charge ratio can be determined. This is particularly helpful in determining isotopic abundance. Mathematically, the centripetal force experienced by the particle with mass m is equal to the Lorentz force. Since we are given that the magnetic field \mathbf{B} is always perpendicular to the particle's velocity (with magnitude v), this gives us $\frac{mv^2}{r} = qvB$, and solving for r gives $r = \frac{mv^2}{qvB} = \frac{mv}{qB}$. The mass spectrometer embodies a simple concept—force depends on mass, so when an equal force is applied to different masses, they accelerate at different rates. The mass spectrometer takes advantage of this by accelerating charged particles in motion using a magnetic field.

Heisenberg's Uncertainty Principle states that it is not possible to simultaneously know a particle's position and velocity with infinite precision on both. Mathematically, $\Delta x \Delta p \geq \hbar/2$. The **Bohr Model** describes the transition energy of an electron between two electronic energy levels (principle quantum numbers):

$$\Delta E = \frac{2\pi^2 m Z^2 e^4}{h^2} \left(\frac{1}{n_{\text{initial}}^2} - \frac{1}{n_{\text{final}}^2} \right) \quad (9)$$

where E is the energy (principle energy level), Z is the nuclear charge, n is the electronic energy level, m is the mass of an electron, e is the charge of the electron, and h is Planck's constant. The important thing is that excitations (i.e. jumping up to higher energy levels) involve an input of energy, while emissions (i.e. releasing energy by going down to lower energy levels) involve the release of energy. Note also that the energies of the electronic energy levels get closer and closer as you go up.

A **paramagnetic species** (referred to as *radicals* in organic chemistry) is defined as an atom or molecule that contains at least one unpaired electron. The extra electron allows the compound to be responsive to an external magnetic field and 'picking up' a magnetic moment, and so can be induced to be magnetic. A **diamagnetic species** has no unpaired electrons, and so are not susceptible to magnetic fields. Here are some important rules for electronic configurations:

- Pauli's Exclusion Principle:** No two electrons can have the same set of quantum numbers.
- Hund's Rule:** Electrons completely fill lower energy levels before starting to fill higher energy levels. In a degenerate set of orbitals, electrons singly occupy each orbital before a second electron pairs up in the same orbital.
- Aufbau Principle:** Electrons are added one-by-one to the shells, starting with the lowest energy level, and then into sequentially increasing energy levels. Exceptions occur in the half-filled and filled d -shells, which add extra stability. For example, Cr (chromium) has $[\text{Ar}]4s^13d^5$ instead of $[\text{Ar}]4s^23d^4$, and Cu (copper) has $[\text{Ar}]4s^13d^{10}$ instead of $[\text{Ar}]4s^23d^9$.

The are four types of quantum numbers to know:

- Principle (n):** Describes the shell and energy level in which the electron resides. $n \in \mathbb{N}$.
- Angular Momentum (l):** Describes the orbital in which the electron resides. l is a whole number less than n . $l = 0$ is an s orbital, $l = 1$ is a p orbital, $l = 2$ is a d orbital, and $l = 3$ is a f orbital.
- Magnetic (m_l):** Describes the orientation of the orbital about a plane or axis. $-l \leq m_l \leq l$.
- Spin (m_s):** Describes the rotation (counterclockwise or clockwise) of the electron about its axis. $m_s = \pm 1/2$.

For p -orbitals ($l = 1$), $m_l = -1$ corresponds to the p_x orbital, $m_l = 0$ corresponds to the p_y orbital, and $m_l = 1$ corresponds to the p_z orbital. For d -orbitals ($l = 2$), $m_l = -2$ corresponds to the d_{xy} orbital, $m_l = -1$ corresponds to the d_{xz} orbital, $m_l = 0$ corresponds to the d_{yz} orbital, $m_l = 1$ corresponds to the $d_{x^2-y^2}$ orbital, and $m_l = 2$ corresponds to the d_{z^2} orbital.

The **coordination number** describes the number of atoms (ligands) attached to a central atom within a molecule. Coordination numbers only count atoms attached, and *not* lone pairs of electrons. For example, BH_3 has a coordination number of 3.

When discussing excitation and relaxation, the generic case is that the energy of the photon absorbed is equal to the energy of the photon emitted. In actuality, there is more than one singular energy level for the ground state and the excited state due to the coupling of electrical energy levels and rotational energy levels associated with the atom. Transitions involve a change in electronic state as well as a change in the rotational energy of the atom. In molecules, there are vibrational energy levels to consider, in addition to the rotational and electronic energy levels. The results is that in actuality, the ground state and excited state exists as a band of energy levels, not a single level. This means that multiple possibilities exist for the energy of transition. Rather than a single line absorption or emission spectrum being formed, a range is formed.

The atomic spectrum of hydrogen and its associated energy transitions are well studied. The **Lyman series** are the transitions that go to the $n = 1$ final energy level. The **Balmer series** are the transitions that go to the $n = 2$ final energy level. The **Paschen series** are the transitions that go the $n = 3$ final energy level. The **Brackett series** are the transitions that go to the $n = 4$ final energy level. Of these four series, only the Balmer series emits photons in the visible range. Here is an example problem:

Given that the wavelength of the emitted light for the $n = 3$ to $n = 2$ transition in the hydrogen atom is 656 nm, which of the following is NOT true?

- A All Balmer series transitions are less than 800 nm.
- B All Lyman series transitions are less than 400 nm.
- C The lowest energy Paschen transition is greater than 800 nm.
- D The highest energy Brackett transition is less than 800 nm.

Let's consider each of the answer choices one at a time:

1. The Balmer series are transitions to the $n = 2$ level, and the lowest energy transition is from $n = 3$ to $n = 2$, and is given to be 656 nm. All other transitions in the Balmer series are more energetic, so these transitions emit photons with a wavelength less than 656 nm. Therefore, choice A is valid.
2. As a general idea, since we know that $\Delta E \propto \left(\frac{1}{n_i^2} - \frac{1}{n_f^2} \right)$, the transition of $n = 2$ to $n = 1$ (which is the lowest energy transition in the Lyman series by definition) as at least three times more energy change than the transition of $n = \infty$ to $n = 2$. Practically, this means that all photons in the Lyman series are of higher energy and shorter wavelength than the transitions in the Balmer series. A more stringent calculation with $\Delta E = hc/\lambda$ would tell us that indeed, the transitions in the Lyman series emit photons with a wavelength less than 400 nm, meaning that choice B is valid.
3. The Paschen series involves transitions to the $n = 3$ level, and the lowest energy transition in the Paschen series is from $n = 4$ to $n = 3$, and it is of significantly lower energy than the $n = 3$ to $n = 2$ transition. This means that the lowest energy photon in the Paschen series is of lower energy and higher frequency than the transitions in the Balmer series. The Balmer series emits a photon at 656 nm, so the lowest energy transition in the paschen series emits a photon with a wavelength greater than 656 nm. Thus, choice C is a valid statement.
4. The Brackett series involves transitions to the $n = 4$ level, and from our above considerations, this should be even less energetic than the Paschen series or Balmer series. Since we showed choice C is valid, the Brackett energy transition cannot be less than 800 nm, so thus choice D is invalid and so choice **D** is the correct answer.

2.3 Periodic Trends

The **effective nuclear charge** Z_{eff} is the net charge exerted upon the valence electrons, which takes into account the shielding from the core electrons. As you move from left to right across a period in the periodic table, the effective nuclear charge increases. Also, as you descend a family/group in the periodic table, the valence shell increases, so the distance between the atom's valence electrons and its nucleus increases. Periodic trends depend on both Z_{eff} and the valence shell. As you move up the periodic table and to the right, the following general atomic trends are observed:

1. The atomic size decreases (the radius of the atom is defined as the distance from the center of the nucleus to the exterior of the valence electron cloud).
2. The ionization energy increases (the energy required to remove the outermost electron from the atom in its gas phase).
3. The electron affinity increases (the energetics associated with an atom gaining an electron in its gas phase).
4. The electronegativity increases (the tendency to hold shared electrons with another atom within a bond).

Note that each trend shows some deviation from uniformity, some more than others. The effective nuclear charge does not uniformly increase when we scan left-to-right across a period. Half-filled stability and filled-shell stability cause some of the common deviations seen with electron affinity and ionization energy. For instance, nitrogen has a larger ionization energy than oxygen, because upon ionization, nitrogen loses its half-filled *p*-shell. On the contrary, oxygen gains half-filled stability upon being ionized.

As a general rule, cations are smaller than neutral atoms because the loss of electrons allows the atom to compact more tightly. As a general rule, anions are larger than neutral atoms, because the gain of electrons causes the atom to expand given the enhanced repulsion associated with the additional electrons.

Helium has a larger **atomic radius** than hydrogen. This is because the electrons in the $n = 1$ level repel one another more than the electrons in any higher quantum level because they have the smallest interelectronic distance. This repulsion forces the electrons away from one another in helium, resulting in a greater area begin occupied by the orbiting electrons, and hence a greater radius for helium.

Ionization energy is the energy required to remove a valence electron when the element is in the gas phase. The generic ionization reaction is represented by the equation $E(g) \longrightarrow E^+(g) + e^-$, where E represents any element. Notable exceptions to the general ionization energy trend occur when there is half-filled stability of the energy level and where there is an s^2 -shell. In both cases, the ionization energy is higher than expected, and also higher than the ionization energy of adjacent elements in terms of atomic number in most cases we'll be considering.

Electron affinity measures the tendency of an element to gain an electron; it can be either negative or positive, meaning that gaining an electron to the valence shell can be either exothermic or endothermic. The generic reaction is represented by the equation $E(g) + e^- \longrightarrow E^-(g)$, where E represents any element. The biggest deviations are attributed to the stability associated with a filled *s*-shell. For instance, the sudden increase in electron affinity from Be to B, Mg to Al, and Ca to Ga is attributed to the instability of one electron in the *p*-level. No trend for electron affinity is evident in the transition metals.

Electronegativity is the ability of an atom to attract towards itself the electrons within a chemical bond. The trend in electronegativity is pretty clean, showing no exceptions at least in the first 20 elements.

2.4 Important Periodic Families

All alkali metals are strong reducing agents and react favorably with water in the reaction $2M(s) + 2H_2O(g) \longrightarrow 2MOH(aq) + H_2(g)$. The oxides they form are variable with the metal. Lithium forms Li_2O , sodium forms a peroxide Na_2O_2 , and potassium, rubidium, and cesium form superoxides MO_2 . Alkali metals can also be oxidized into cations by halogens, nitrogen, and hydrogen.

Alkaline earth metals are not as soluble in water as are the alkali metals in their cation form, primarily due to their 2^+ charge and smaller radius. They can react favorably with water (except for Be) according to the reaction $M(s) + 2H_2O(g) \longrightarrow M(OH)_2(aq) + H_2(g)$. The alkaline earth metals all form oxides MO when oxidized by oxygen gas. They can also be oxidized by halogens, nitrogen, and hydrogen.

Chalcogens are the sixth group of the periodic table that includes oxygen, sulfur, selenium, tellurium, and polonium. As neutral elements, they are oxidizing agents, but their reactivity decreases as you descend the column. Oxygen exists as a diatomic molecule O_2 , sulfur and selenium exist as octatomic molecules S_8 and Se_8 , and tellurium and polonium exist in vast molecular matrices.

Halogens and noble gases are also important but nothing's new here. When exposed to high concentrations of fluorine gas, xenon and krypton are known to form molecular compounds with fluorine. The resulting xenon fluorides and krypton fluorides show more electron density around fluorine than the noble gas, implying that fluorine is more electronegative than either xenon or krypton. Otherwise, noble gases are pretty much unreactive.

2.5 Electrochemistry

Electrochemistry addresses chemical reactions that involve electron transfer. **Galvanizing** is a protective process that involves plating a more reactive metal onto a less reactive metal, so that this more reactive species (i.e. the *sacrificial metal*) is the one oxidized by air, rather than the material being protected.

Generally speaking, organic and biological oxidizing agents are rich in oxygen and poor in hydrogen (e.g. Na_2CrO_4 , KMnO_4 , NAD^+ , and FAD), and reducing agents are poor in oxygen and rich in hydrogen (e.g. NaBH_4 , LiAlH_4 , NADH , and FADH_2). This is because oxidizing agents get reduced while oxidizing whatever they're reacting with, and reducing agents get oxidized while reducing whatever they're reacting with.

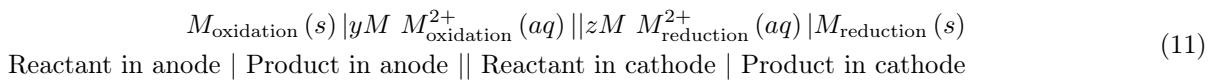
The reduction of two protons (H^+) to form hydrogen gas (H_2) is defined as the reference standard, and assigned an *emf* of zero volts. Any compound that can be reduced more favorably than a proton (meaning oxidation is less favorable than a proton) has a positive reduction potential. Likewise, any compound for which reduction is less favorable than a proton (meaning oxidation is more favorable than a proton) has a negative reduction potential. Generally speaking, precious metals do not readily oxidize—thus, they have relatively high reduction potentials. On the other end of the spectrum are the alkali and alkaline earth metals, which have relatively low ionization energies, so they are easily oxidized, meaning they have negative reduction potentials. Most tables list reduction half-reaction *emf* values, rather than oxidation half-reaction values. *Emf* values do not need to be memorized, but in general, the greater the electron affinity, the greater the reduction potential. Likewise, a lower ionization energy corresponds to a greater oxidation potential (lower reduction potential).

The free energy in an electrochemical cell can be described by

$$\Delta G_{\text{reaction}} = -nF\varepsilon_{\text{cell}} \quad (10)$$

where $F = 96500$ C per mole and n = electrons per reaction. A positive electromotive force (cell voltage) is associated with a favorable oxidation-reduction reaction, just as a negative free energy change (ΔG) is associated with a favorable oxidation-reduction reaction.

Electrochemical cells convert energy produced in a chemical reaction into electrical current. Oxidation occurs at the anode, meaning that electrons flow away from the anode, and reduction occurs at the cathode (REDCAT), meaning that electrons flow towards the cathode. This means that anions migrate towards the anode and cations migrate towards the cathode through electric fields. The important thing to remember is that **electrons flow from the anode to the cathode**. There is a standard line notation for cells—when reading left to right, it goes from anode to cathode, and reactant to product within each half-reaction electrode. For example,



In general, there are two types of electrochemical cells: **galvanic cells** (spontaneous cell with $\varepsilon^\circ > 0$, $\Delta G < 0$) and **electrolytic cells** (non-spontaneous cell with $\varepsilon^\circ < 0$, $\Delta G > 0$, and an applied voltage present to power the cell). Galvanic cells release energy in the form of electrical flow, while electrolytic cells are used for the storage of electrical potential, as is seen when recharging a battery.

The **Nernst equation** is often used to quantify the effect of concentration on the voltage of an electrochemical cell. Mathematically, it is

$$\varepsilon_{\text{cell}} = \varepsilon^\circ - \frac{RT}{nF} \ln Q_{rx} \quad (12)$$

By substituting $R = 8.314 \text{ J/mol} \cdot \text{K}$, $T = 298 \text{ K}$, and $F = 96500 \text{ C/mol}$, we get the working equation

$$\varepsilon_{\text{cell}} = \varepsilon^\circ - \frac{0.059}{n} \log_{10} Q_{rx} \quad (13)$$

where we're being pretty lenient with the units. Note that our working equation has \log_{10} instead of \ln natural logarithm. Here is an example problem using the Nernst equation:

What is the voltage of an electrochemical cell with an anode of zinc metal in 0.01 M $\text{Zn}(\text{NO}_3)_2$ (aq) and a cathode of silver metal in 1.00 M AgNO_3 (aq)?

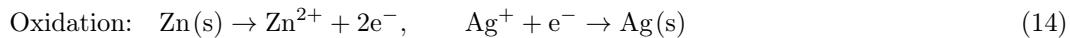
A 2.42 V

B 2.30 V

C 1.62 V

D 1.50 V

First, we must identify the half-reaction:



The standard cell potential for the reaction is found using values from online or a provided table (not the important part here), and we get $\varepsilon^\circ = 0.80 \text{ V} - (-0.76 \text{ V}) = 1.56 \text{ V}$. The actual cell voltage is slightly higher than 1.56, because there is a higher cation concentration in the cathode than in the anode. The exact value is found using the Nernst equation:

$$\varepsilon_{\text{cell}} = \varepsilon^\circ - \frac{0.059}{n} \log \frac{[\text{Zn}_{\text{anode}}^{2+}]}{[\text{Ag}_{\text{cathode}}^+]^2} = 1.56 \text{ V} - \frac{0.059}{2} \log \frac{0.01}{1^2} = (1.56 - 0.03(\log 0.01)) \text{ V} = (1.56 + 0.06) \text{ V} = 1.62 \text{ V} \quad (15)$$

giving us the answer choice C as the correct answer, as desired.

2.6 Gases and Gas Laws

The **collision force** is defined as the force exerted by a gas particle during a collision between it and the container wall. It is based on impulse, so both greater momentum and a shorter time of contact increase the force of impact. The collision force can be increased by increasing the temperature, because greater temperature imparts greater velocity, and thus greater momentum, to each molecule. The **mean free path** is defined as the average distance a particle can travel before colliding with another particle. There are three key assumptions associated with the **Kinetic Molecular Theory of Gases**:

1. Particles are so small compared to the distances between particles that their volumes are negligible.
2. Particles move in straight lines, and there are no intermolecular forces. The collisions are said to be elastic with no energy dissipation.
3. Particles are in constant random translational motion.

An ideal gas is a theoretical gas that obeys all of these conditions. In a gas system, not all particles have the same kinetic energy—there is a random distribution of energies defined by the temperature-dependent *Boltzmann distribution*. Of course, no gas is truly ideal, and we have the **van der Waals equation** to describe real gases:

$$\left(P_{\text{observed}} + a \frac{n^2}{V^2} \right) (V_{\text{container}} - nb) = nRT \quad (16)$$

where a is the **attraction coefficient** ($a > 0$ is particles attracting, $a < 0$ is particles repelling), b is the '**bigness coefficient**' ($b > 0$ for all particles). Here are two sample problems regarding the van der Waals equation:

Which of the following types of gases has a negative a term?

A Polar

B Nonpolar

C Hydrophobic

D Ionized

An ionized gas has particles that all have the same charge, and like charges repel, so choice D is the best answer.

Which of the following types of gases has the greatest b term?

- A Methane
- B Ethane
- C Ethene
- D Ethyne

Ethane has the greatest mass of the given molecules, and so it has the greatest b term. Thus, the answer is **B**.

Here's another practice problem, this time regarding how to measure pressure using a manometer and using the equation $\Delta P = \rho g \Delta h$:

What is the pressure in atmospheres of a column of gas in a closed tube above mercury if the height difference at sea level between a connected column of mercury open to the atmosphere and the closed column above mercury is 317 mm?

- A 317/760 atm
- B 443/760 atm
- C 760/443 atm
- D 317/443 atm

The height difference of 317 mm means that the pressure difference is 317 torr (mm Hg, by definition). The atmospheric pressure at sea level is 760 torr, so the gas pressure in the column is $760 - 317 = 443$ torr. The conversion from torr to atm involves dividing by a factor of 760 torr per atm, making choice **B** the best answer. As the question is worded, the pressure difference is provided, but the relative pressure are not mentioned. The pressure could also be $760 + 317 = 1077$ torr, but this value is not listed as an answer choice.

The **root mean square speed** of a gas is

$$\mu_{\text{rms}} = \sqrt{\frac{3RT}{m}} \quad (17)$$

where R is the ideal gas constant, T is the temperature, and m is the mass of one mole of the gas. This is derived by solving the equation $\frac{1}{2}mv^2 = \frac{3}{2}RT$ for v , as the left hand side is the kinetic energy from classical mechanics and the right hand side is the energy of the system obtained through the equipartition theorem. Here is an example problem:

What is the root mean square speed of neon atoms at 27°C?

- A 19.3 m/s
- B 211 m/s
- C 612 m/s
- D 1018 m/s

From the periodic table, the mass of one mole of neon is $m = 0.20$ kg, and the temperature converted to Kelvin is $T = 300$ K. Thus,

$$\mu_{\text{rms}} = \sqrt{\frac{3RT}{m}} = \sqrt{\frac{3(8.314)(300)}{0.020}} = \sqrt{\frac{(8.314)(9 \times 10^2)}{2 \times 10^{-2}}} \approx \sqrt{\frac{8 \times 9}{2} \times 10^4} = \sqrt{36 \times 10^4} = 6 \times 10^2 = 600 \text{ m/s} \quad (18)$$

where we ignore the units for convenience. Thus, the best answer is choice **C**. Using this formula for the root mean square speed, we can determine the relative speeds of two gases, giving us **Graham's law** for gas flow:

$$\frac{v_2}{v_1} = \sqrt{\frac{m_1}{m_2}} \quad (19)$$

which holds for a given temperature. This equation is basically saying that the speed of the particle is inversely proportional to the square root of its mass, so the lighter the gas, the faster it travels at a given temperature. There is also the relationship

$$\frac{v_2}{v_1} = \sqrt{\frac{T_2}{T_1}} \quad (20)$$

which holds for a given mass of particles. This is basically saying that the speed of the particle is directly proportional to the square root of the temperature, so the greater the temperature, the faster a gas travels.

2.7 Phases and Phase Changes

Isochoric conditions are where the *volume* of the system does not change, and **adiabatic** conditions are where *heat* neither enters nor exits the system, meaning that the system is perfectly insulated. The **critical point** is the highest temperature and pressure at which a liquid may be observed. Beyond the critical point, it is impossible to distinguish between a gas and a liquid, so it is referred to as a *supercritical fluid*. The **triple point** is where all three phases of matter can exist for a given compound. It is also the highest temperature and pressure at which sublimation (and deposition) can occur. Two key facts of water are that its liquid form is denser than its solid form, and it is densest at 4°C.

The enthalpy of vaporization is greater in magnitude than the enthalpy of fusion because more energy is necessary to *break* the intermolecular forces (i.e. for liquid → gas) than is necessary to weaken the intermolecular forces (i.e. for solid → liquid).

The formal definition of **vapor pressure** above a liquid is the force per unit area above the surface of a liquid exerted by molecules formed upon evaporation of the liquid. Vapor pressure can be measured in either an open or closed system. In a closed system, a partial pressure of vapor exists, because the rate of vaporization equals the rate of condensation. This is most typically how vapor pressure is determined. In an open system, since no equilibrium is reached, vapor pressure is a measure of the pressure just above the surface of the liquid. Vapor pressure is *independent* of the shape and volume of a container. Vapor pressure increases as temperature increases, and decreases as the heat of vaporization $\Delta H_{\text{vaporization}}$ increases. The **Clausius-Clapeyron** equation can be used to explain the relationship between vapor pressure and temperature:

$$\ln \frac{p_1}{p_2} = \frac{\Delta H_{\text{vaporization}}}{R} \left(\frac{1}{T_2} - \frac{1}{T_1} \right) \quad (21)$$

The conclusion from this is that as the temperature of a liquid increases, the pressure of its vapor increases in an exponential manner. This equation assumes that the heat of vaporization $\Delta H_{\text{vaporization}}$ is constant over the temperature range of interest. Detailed calculations involving the Clausius-Clapeyron equation are unlikely, but it should still be understood conceptually and graphically.

$$\text{As } \Delta H_{\text{vaporization}} \uparrow; \text{ Volatility } \downarrow; \text{ Boiling Point } \uparrow; \text{ Vapor Pressure } \downarrow \quad (22)$$

The **boiling point** is defined as the temperature at which the vapor pressure is equal to the atmospheric pressure. According to **Raoult's Law**, the vapor pressure above a solution of two or more *miscible* liquids depends on the mole fraction of each compound in solution. Mathematically,

$$P_{\text{vapor}} = \xi_i P_{\text{vapor (pure)}} \quad (23)$$

where the ξ_i in Raoult's equation is the mole fraction of the component *in solution*, not in the vapor state. With real compounds, because of intermolecular forces within liquids, there are other factors that come into play. The molecules interact in both the liquid and gas phases, so the vapor pressure relationship is not *exactly* linear as Raoult's equation approximates it to be. Intermolecular forces are greater in the solution than the gas phase. If there is an overall increase in attractive forces in solution when the components are mixed, then vaporization, and thus vapor pressure, decreases, resulting in a negative deviation from linearity. Likewise, if there is an overall decrease in attractive forces in solution when the components are mixed, then vaporization, and thus vapor pressure, increases, resulting in a positive deviation from linearity. **Distillation** is removing a liquid from a solution by utilizing their differing vapor pressures/boiling points.

2.8 Colligative Properties

Colligative properties are properties of a solution that are affected by the concentration of a soluble impurity. Colligative properties include boiling point elevation, freezing point depression, and osmotic pressure. The electrical conductivity of a

solution, although it is not formally a colligative property, shows a dependence on ion concentration. The increase in boiling point ΔT_b when solute is added to a solution is given by

$$\Delta T_b = k_b \cdot i \cdot m \quad (24)$$

where i is the number of ions that form upon dissolving, m is the *molality* of the solution, and k_b is a constant for the solvent. Similarly, we also have the equation for freezing point depression which basically looks the same as that for boiling point elevation:

$$\Delta T_f = k_f \cdot i \cdot m \quad (25)$$

Note that $\Delta T_f < 0$ and $\Delta T_b > 0$ (that's why it's called freezing point *depression* and boiling point *elevation*).

Water has a natural tendency to flow from solutions of higher water concentration (i.e. lower solute concentration) to solutions of lower water concentration (i.e. higher solute concentration) to reach equal concentrations. Pressure differences cause fluids like water to flow, and the driving force causing water to flow is known as **osmotic pressure**. Osmotic pressure π is the pressure exerted by a solution through osmosis across a semipermeable membrane, and is given by

$$\pi = MiRT \quad (26)$$

where M is the *molarity*, i is the number of ions upon dissociation, R is the ideal gas constant, and T is the temperature. The **condosity** of a solution is defined as the molar concentration of an aqueous sodium chloride solution that has the same specific conductance as the aqueous salt solution.

2.9 Solubility

A **supersaturated** solution is where the amount of solute that is dissolved into solution is beyond the maximum amount at that given temperature. The solution is actually a suspension that when disturbed can form a precipitate rapidly. This state can be achieved by first heating a solvent, then adding solute to the solution until the solution is saturated at that temperature. Slowly cooling this solution causes the amount of solute in it to exceed that which should dissolve at the reduced temperature. **Gram solubility** is the quantitative measurement of the maximum number of grams of solute that can dissolve into enough solvent to make one hundred mL of solution under ambient conditions. Here are the solubility rules you should know:

1. Most salts composed of +1 cation (excluding transition metals) and -1 anion are soluble in water at room temperature. Nitrate NO_3^- salts are water-soluble.
2. Most salts containing sulfate SO_4^{2-} anions *with +1 cations* (excluding transition metals) are water-soluble.
3. Most salts with -2 or -3 anions are insoluble in water (excluding sulfate salts).
4. Most oxide O^{2-} and hydroxide OH^- anion salts are only slightly water soluble. KOH and NaOH are notable exceptions that are substantially water soluble.

An **ion-exchange** column exchanges one ion in solution for another (which is initially bound to the column), by precipitating out the ion that forms the less soluble salt and releasing a more soluble into solution to replace it. A water softener is a good example of an ion-exchange column—hard water (rich in $\text{Ca}^{2+}(\text{aq})$) travels down the column where the anion in the ion-exchange resin binds calcium cations to form an insoluble salt while releasing sodium cations $\text{Na}^+(\text{aq})$ into solution. Aqueous sodium cation is referred to as soft water. As calcium precipitates through the column, it is filtered from the water, preventing it from forming precipitates in the plumbing lines of appliances. The reaction that we're interested in is $2\text{Na}^+(\text{aq}) + \text{CaCl}_2(\text{s}) \longleftrightarrow \text{Ca}^{2+}(\text{aq}) + 2\text{NaCl}(\text{s})$; this reaction highly favors the reactants side. There's also another thing called **selective precipitation and ion exchange**, where we add an ion to solution with the intention of precipitating an existing ion out of the solution. This is basically the same thing as the ion-exchange column, and can be achieved when the salt being precipitated is less soluble than any other salt combination in solution. For example, consider the reaction $2\text{AgCl}(\text{s}) + \text{Hg}^{2+}(\text{aq}) \longleftrightarrow 2\text{Ag}^+(\text{aq}) + \text{HgCl}_2(\text{s})$. Because mercury chloride is less soluble than silver chloride, the addition of silver chloride to the mercury solution precipitates mercury out of water.

When reactions are added together, the equilibrium constant K for the overall reaction is the product of the equilibrium constants of the component reactions. Here is an example problem for solubility:

Silver chloride AgCl is MOST soluble in which of the following solutions at 25°C?

- A 0.10 M HgNO₃(aq)
- B Pure water
- C AgNO₃(aq)
- D NaCl

As discussed above, silver chloride is most soluble in a solution where Hg⁺(aq) is present, as the mercury ion in solution can bind the chloride anion to form a precipitate. As the amount of Cl⁻ in solution decreases as a result, more AgCl dissolves into solution to replace it. Thus, the best answer is choice A. Both choices C and D are eliminated because of the common ion effect.

2.10 Acids and Bases

The **haloacids** are the series of HX acids, where X represents a halogen. As the size of the halogen increases, the haloacid becomes more acidic due to the increased stability of the conjugate base halide as it increases in size. Dissociation increases, bond strength decreases, and bond length increases as X gets bigger.

Within a period of the periodic table, it is electronegativity that dictates the strength of an acid, *not* atomic radius. A prime example of this idea is the relationship between ammonia H–NH₂, water H–OH, and hydrofluoric acid H–F. The strongest acid of the three is HF, because fluorine is more electronegative than oxygen or nitrogen. The atomic size doesn't change that noticeably between the three elements—the periodic trend that changes the most from N to F is the electronegativity.

To determine how much acid (or base) is necessary to neutralize a solution of base (or acid), use $M_1V_1 = M_2V_2$. Make sure to take into account if the acid is polyprotic/the base has multiple hydroxide groups or something like that. Here is an example problem:

Citric acid (C₆H₈O₇) has three dissociable protons with pK_a values of 3.14, 4.79, and 5.20. Which of the following solutions would have the lowest pH value?

- A 50 mL 0.10 M citric acid (aq) + 75 mL 0.20 M NaOH(aq)
- B 25 mL 0.10 M citric acid (aq) + 50 mL 0.10 M NaOH(aq)
- C 25 mL 0.20 M citric acid (aq) + 75 mL 0.20 M NaOH(aq)
- D 50 mL 0.20 M citric acid (aq) + 50 mL 0.10 M NaOH(aq)

The lowest pH belongs to the solution that is most acidic, and the most acidic solution is the one where the fewest equivalents of base relative to citric acid have been added. In choice D, the NaOH(aq) solution is of equal volume and half the concentration of the citric acid solution, so there is only one-half of an equivalent of NaOH(aq). Comparing this with the other answer choices, this is the lowest number of equivalents of the base compared to the other answer choices, so thus, choice D is the best answer.

In a solution, the distribution within a conjugate pair is dictated by the pH of the solution. The conjugate pair favors the conjugate acid form in the presence of hydronium, and favors the conjugate base form in the presence of hydroxide. In general, if pH>pK_a, the solution is basic relative to the compound, so it exists predominantly in its deprotonated form ([A⁻]>[HA]). If pH<pK_a, the solution is acidic relative to the compound, so it exists predominantly in its protonated form ([A⁻]<[HA]).

What is the pK_a for ammonia NH_3 , given that the pK_b for ammonia is 4.7?

- A 4.7
- B 7.0
- C 9.3
- D 33

It is not 9.3! The compound with a pK_a of 9.3 is ammonium (NH_4^+), the conjugate acid of ammonia. Ammonia is a weaker acid than ammonium, so it has a pK_a greater than 9.3. The best answer is 33, choice D, indicating that ammonia is such a weak acid that when added to water, there is no detectable dissociation. The mistake of choosing 9.3 is easy to make, one that most students make routinely. However, the equation $pK_a, (\text{HA}) + pK_b, (\text{A}^-) = 14$ is for conjugate pairs, not for the same compound. The pH of a solution composed of both components in a weak conjugate pair can be determined using the **Henderson-Hasselbalch** equation:

$$\text{pH} = \text{p}K_a + \log \frac{[\text{HA}]}{[\text{A}^-]} \quad (27)$$

The equation shows that as the conjugate base concentration [HA] increases, the pH of the buffer increases.

2.11 Titration Curves

A **buffer** is a solution whose pH remains relatively constant after the addition of either strong acid or strong base. The pH may vary slightly, but for all intents and purposes, it does not change significantly. The pH of a buffered solution is determined using the Henderson-Hasselbalch equation, as discussed in the previous section. Buffers are made of a roughly equal mole mixture of a weak acid and its weak conjugate base in an aqueous solution. The pH range of a buffer is $\text{pH} \in [\text{p}K_a - 1, \text{p}K_a + 1]$. Buffers can be made by mixing any of the following things together:

1. Weak acid + the salt of the conjugate base in roughly equal mole proportions (e.g. HCO_2H with HCO_2Na).
2. Weak base + the salt of the conjugate acid in roughly equal mole proportions (e.g. NH_3 with NH_4Cl).
3. Weak acid + roughly half of an equivalent of strong base (e.g. HOAc with a half-equivalent of KOH).
4. Weak base + roughly half of an equivalent of strong acid (e.g. H_3CNH_2 with a half-equivalent of HCl).

Physiological pH is considered to be 7.4 and is usually highly buffered in most regions of the body. In a *titration*, the **equivalence point** is reached when the moles of acid are equal to the moles of base. Here are a couple of good practice problems for titrations:

What is the pH after 30 mL of 1.00 M NaOH (aq) have been added to 100 mL 0.50 M HOAc (aq)? HOAc has a $\text{p}K_a=4.74$.

- A 3.51
- B 4.56
- C 4.92
- D 5.97

Because the strong base is twice as concentrated as the weak acid, only half the volume of strong base (relative to the weak acid) is required to reach the equivalence point. This means that 50 mL of 1.00 M NaOH fully neutralizes the 100 mL of 0.50 M HOAc . The halfway point of the titration is therefore reached when exactly 25 mL of 1.00 M NaOH have been added. At the halfway point, the pH of the solution equals the $\text{p}K_a$ of the weak acid. The additional strong base beyond the 25 mL makes the pH of the solution slightly greater than the $\text{p}K_a$ of the acid, 4.74. Thus, the best choice is answer C. Here is another problem:

What is the pH of a solution made by mixing 10.0 mL 0.10 M HCO_2H (aq) with 4.0 mL 0.10 M KOH(aq)? The pK_a of HCO_2H is 3.64.

- A 1.34
- B 3.46
- C 3.82
- D 9.36

The weak acid and titrant strong base are of equal concentration, so 10 mL KOH is one equivalent (the amount needed to reach the equivalence point). If 5.0 mL are added, then the acid is half-titrated, so $\text{pH} = \text{pK}_a$. However, less than 5.0 mL have been added, so the pH is a little less than pK_a . The pK_a is given to be 3.64, so the best answer is choice B. Choice A is too much less than the pK_a .

Polyprotic acids yield multiple equivalents of hydronium ion per molecule, and so therefore have multiple equivalence points, one of each dissociable proton. Each equivalence point is the arithmetic mean of the two adjacent pK_a values. This is clearly shown in Fig. 2.11.

A typical acid-base **indicator** is an organic acid with extended conjugation. Its weak acid and conjugate weak base

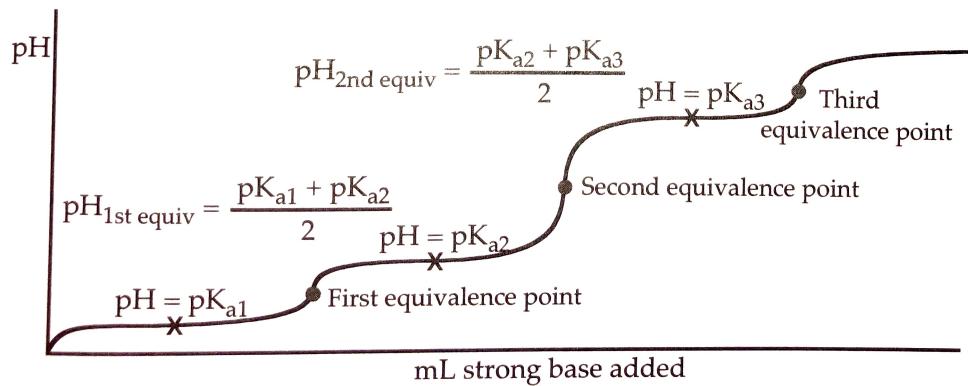


Figure 25: An example of a titration curve with a polyprotic acid.

have two distinct colors. The energy state transition that produces color involves the π -bonding and π -antibonding orbitals, and depends on the electron-donating and electron-withdrawing nature of the substituents on the π -system. An indicator is generally used for one of two purposes: the first is to detect the endpoint of a titration, and the second is to approximate the pH of a solution from the color of the indicator in the solution. The best scenario for an indicator detecting a titration endpoint is when the pH at equivalence equals the pK_a of the indicator. The important fact to retain is that the pH at equivalence for the titration of a weak acid by a strong base can be approximated quite closely by averaging the pK_a of the weak acid with the pH of the titrant strong base. For the titration of a weak base by a strong acid, the pH at equivalence can also be approximated by averaging the pK_a of the conjugate acid of the weak base with the pH of the titrant strong acid.

2.12 Equilibrium

The **equilibrium constant** K_{eq} is a mathematical quantity that is calculated for a reaction at equilibrium. By definition, the equilibrium constant is the concentration of products at equilibrium divided by the concentration of reactants at equilibrium. Stoichiometric values from the balanced equation become exponents in the K_{eq} expression. Do *not* include solids or pure liquids in the K_{eq} expression, only solutes (for K_c) and gases (for K_p). The numerical value of K_{eq} varies only with changing temperature, not with catalysts, pressure, volume, or moles. Equilibrium constants are unitless by definition. All equilibrium constants obey the same rules, but depending on the reaction, there may be special features that recur. Different reactions have special K values.

A **complex equilibrium** is a balance between two separate reactions that share a common reagent or set of reagents. In essence, it is two reactions whose equilibrium states depend on one another. Here is an example of a complex equilibrium:

| K Type | Type of Reaction to Which the K Applies |
|----------|---|
| K_p | K_{eq} for reaction of gases. Values are in <i>pressure</i> units. |
| K_c | K_{eq} for reaction of solutes. Values are in <i>concentration</i> units. |
| K_{sp} | K_{eq} for salts dissociating into ions. Measures the <i>solubility</i> . |
| K_a | K_{eq} for acids dissociating into water. Measures the <i>acidity</i> . |
| K_b | K_{eq} for bases dissociating in water. Measures the <i>basicity</i> . |
| K_w | K_{eq} for autoionization of <i>water</i> into H^+ and OH^- . |

Consider the following complex equilibrium. When N_2O_4 gas is removed, how are the partial pressures of NO gas and NO_2 gas affected?



(28)



- A P_{NO} and P_{NO_2} both decrease.
- B P_{NO} and P_{NO_2} both increase.
- C P_{NO} decreases, while P_{NO_2} remains the same.
- D P_{NO} increases, while P_{NO_2} remains the same.

The second reaction shifts to the right to compensate for the loss of N_2O_4 , so the partial pressure of NO_2 decreases. In turn, a decrease in NO_2 causes the first reaction to shift to the right as well, resulting in a decrease in the partial pressure of NO gas and an increase in the partial pressure of NO_2 . However, this increase in NO_2 from the forward shift of the first reaction is less significant than the decrease in the NO_2 caused by the forward shift of the second reaction. This is because the shift in the first reaction cannot completely replenish the lost NO_2 without losing so much NO that the reaction is beyond equilibrium. A shift *never* regenerates as much as was lost. Both NO and NO_2 thus decrease, making choice A the best answer.

Equilibrium is **dynamic**, meaning that it is a state where the system is continually reacting in both the forward and reverse directions, even though no net change is observed at the macroscopic level. **Le Châtelier's Principle** says that if an external stress is applied to a system at equilibrium, the system will shift itself in such a way that the stress is partially relieved as best as possible and equilibrium is re-established. To see this in action, consider the reaction $\text{N}_2\text{O}_4 \rightleftharpoons 2 \text{NO}_2(\text{g})$, which is an endothermic reaction with $\Delta H = +146 \text{ kJ/mole}$:

1. If $\text{N}_2\text{O}_4(\text{g})$ is added, the system shifts to the right.
2. If $\text{NO}_2(\text{g})$ is added, the system shifts to the left.
3. If the external pressure is increased, the system favors the side with less molecules to reduce the increased crowding, thus shifting the system to the left.
4. If the volume is increased, the system is less crowded and molecules collide less often, so the system shifts to the right.
5. If the system is heated, this gives a net influx of energy, shifting the *endothermic* system to the right.
6. If the system is cooled, the system wants to produce energy to compensate, shifting the system to the left.

2.13 Thermochimistry

Exergonic means $\Delta G < 0$, while *exothermic* means $\Delta H < 0$. Similarly, *Endogonic* means $\Delta G > 0$, while *endothermic* means $\Delta H > 0$. The standard free energy change ΔG° is the free energy change when a reaction goes from standard conditions

to equilibrium. Standard conditions are defined as 1 atm of pressure and a temperature of 298 K, and all solute species (reactants and products) present at 1.00 M concentration. We have

$$\Delta G^\circ = -RT \ln K_{eq} \quad (29)$$

If we don't start at reaction quotient $Q = 1$ by not having all the solute species at 1.00 M concentration, then the adjusted free energy change is

$$\Delta G_{rxn} = RT \ln \frac{Q_{rxn}}{K_{eq}} \quad (30)$$

2.14 Chemical Kinetics

The **order** of a reaction is ascertained from the number of reagents involved in the rate-determining step. There are three fundamental reaction orders to consider: zero, first, and second. The rate of a reaction is affected by several factors such

| | Zero Order | First Order | Second Order |
|-------------------|---|---|---|
| Rate Law: | $\text{Rate} = k$ | $\text{Rate} = k[A]$ | $\text{Rate} = k[A]^2$ |
| Half-Life: | $t_{1/2} = \frac{[A]_0}{2k}$ | $t_{1/2} = \frac{\ln 2}{k} = \frac{0.693}{k}$ | $t_{1/2} = \frac{1}{k[A]_0}$ |
| Rate: | Constant rate | Rate \downarrow as time \uparrow | Rate \downarrow as time \uparrow |
| Half-Life: | Half-life \downarrow as time \uparrow | Half-life constant | Half-life \downarrow as time \downarrow |

as temperature, activation energy, catalysts, solvent (solvents affect the stability of the transition state), collision frequency, collision orientation, and the concentration of the reactants in the rate-determining step. The rate constant k must account for all of these factors, except concentration of the reactants, and is given by

$$k = A \exp\left(\frac{-E_{\text{activation}}}{RT}\right) \quad (31)$$

where A is the Arrhenius constant—it takes into account collision orientation and frequency. Because half-life is constant for first-order processes, most half-life questions on the MCAT involve decay (a first-order reaction), which can be described by

$$C_t = C_0 e^{-kt} \quad (32)$$

where C_t is the concentration at time t , C_0 is the initial concentration, t is time, and k is the rate constant. The half-life of decay is $t_{1/2} = (\ln 2)/k \approx 0.693/k$.

In nuclear chemistry, a heavy element undergoes a decay process (i.e. fission) to increase its nuclear stability. A light element undergoes a capture process (i.e. fusion) when struck by a high-energy particle to increase its nuclear stability.

3 Organic Chemistry of Biological Systems

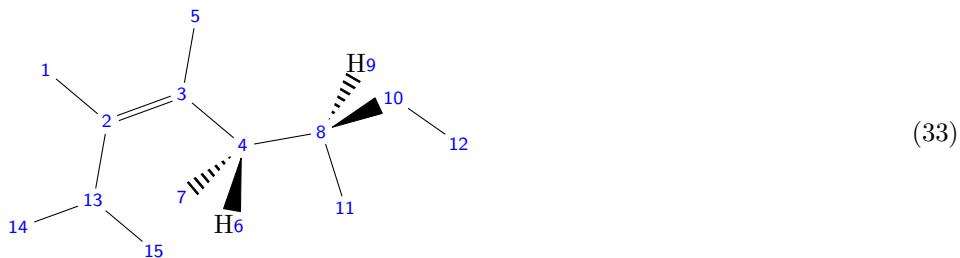
3.1 Molecular Structure

For nonpolar σ bonds, most of the electron density is *between* the nuclei. This means that when drawing electron cloud diagrams, the σ bond looks more like an ellipsoid as opposed to a dumbbell.

The more substituted a carbon atom is, the less strong bonds it is able to form. This is because the other groups are able to (mildly) take electrons away from the bond of interest and thus make it weaker. Bonds involving sp^3 carbons are the weakest, while those involving sp carbons are the strongest, because they have the most s character. And of course, triple bonds are the strongest, while single bonds are the weakest, because of the number of electrons shared between atoms. For

| Process | Reaction Tracking |
|----------------------|---|
| α -Decay | $^{120}_{50}\text{X} \longrightarrow {}_2^4\alpha + {}_{48}^{116}\text{Y}$ |
| β -Decay | $^{120}_{50}\text{X} \longrightarrow {}_{-1}^0\text{e} + {}_{51}^{120}\text{Z}$ |
| β^+ -Decay | $^{120}_{50}\text{X} \longrightarrow {}_1^0\text{e} + {}_{49}^{120}\text{Q}$ |
| γ -Emission | $^{120}_{50}\text{X}^* \longrightarrow h\nu + {}_{50}^{120}\text{X}$ |
| α -Capture | $^{120}_{50}\text{X} + {}_2^4\alpha \longrightarrow {}_{52}^{124}\text{A}$ |
| β -Capture | $^{120}_{50}\text{X} + {}_{-1}^0\text{e} \longrightarrow {}_{49}^{120}\text{Q}$ |
| β^+ -Capture | $^{120}_{50}\text{X} + {}_1^0\text{e} \longrightarrow {}_{51}^{120}\text{Z}$ |
| γ -Absorption | $^{120}_{50}\text{X} + h\nu \longrightarrow {}_{50}^{120}\text{X}^*$ |

example, consider the molecule shown in Figure 33.



Here, let us consider Bonds A (1 and 2), B (2 and 3), C (2 and 13), and D (4 and 8). Between these four bonds, B is the strongest since it is a double bond. D is the weakest because it is between two sp^3 -hybridized carbons. Bond A is stronger than bond C, despite both sharing an sp^2 -hybridized and sp^3 hybridized carbon, because bond C contains the more highly substituted carbon.

Hybridization of carbon atoms can affect the acidity of hydrogens bound to them. Contrary to most other acids, as the hybrid orbital gets smaller, the electrons are held closer to the nucleus of the atom bonded to hydrogen, so the bond can be cleaved in a heterolytic fashion more easily. The more *s*-character in the hybrid orbital of the atom bonded to the hydrogen, the stronger the acid. This results in the relative acidity being $sp > sp^2 > sp^3$. This trend is most commonly observed with C acidity, but can also be observed with N and O.

Steric hindrance occurs any time two atoms attempt to be in the same place at the same time. It is repulsive in nature and increases as the atoms draw closer. On cyclohexane, substituents with axial orientation experience greater steric hindrance than substituents with equatorial orientation. Thus, substituents (especially the bulky ones) prefer to be in the equatorial position.

Intermolecular forces are the primary consideration when approximating physical properties (i.e. boiling point). If IMFs are not enough, you also need to consider the molecular mass and molecular rigidity of the molecules. Heavier compounds have higher boiling points, and compounds with greater molecular flexibility can twist and conform to allow for more surface area, and thus more intermolecular interactions. This is why cell membranes would be most rigid if their fatty acids were completely saturated and long molecules, as interactions are greatest with long, saturated fatty acids.

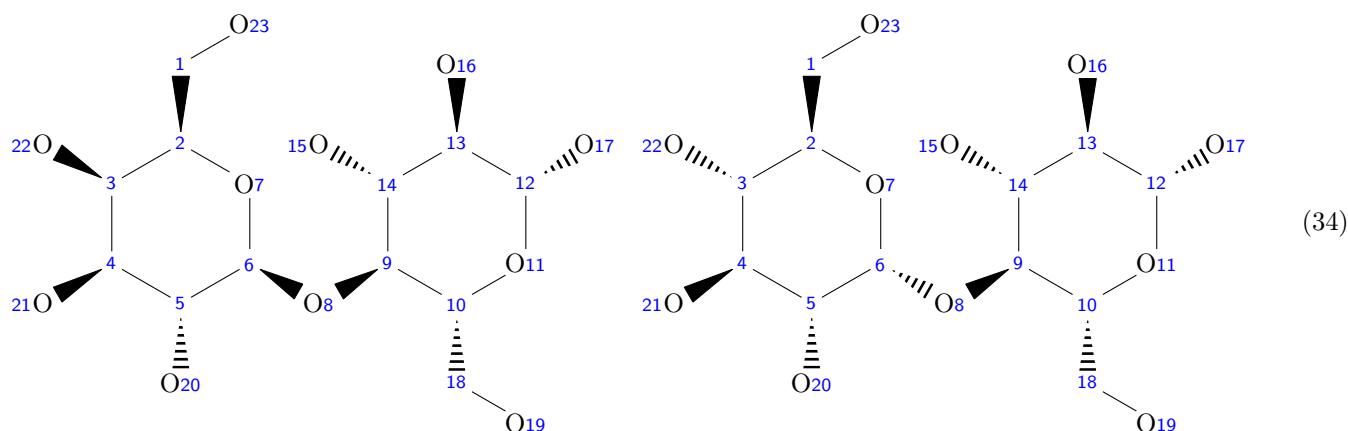
The best micelles has an ionic (charged) head and a long carbon chain for the organic tail. For example, $\text{H}_3\text{C}(\text{CH}_2)_{14}\text{CO}_2^-$ is a better micelle than $\text{H}_3\text{C}(\text{CH}_2)_{14}\text{CO}_2\text{H}$ because it has a charged head, which is more hydrophilic than even the polar, protic carboxylic acid.

The inductive effect can be applied to both electron withdrawal (i.e. with the canonical example of nearby halogens) and electron donation. For instance, methyl amine is more nucleophilic than ammonia (NH_3) because the methyl group is electron-donating. Varying the R-group changes the inductive effect. It also changes the size of the molecule, so steric hindrance can affect the reaction. For instance, $(\text{H}_3\text{C})_3\text{N}$ is less nucleophilic than $(\text{H}_3\text{C})_2\text{NH}$ because the electron donation by the additional methyl group does not compensate for the increase in molecular size.

Stereochemistry prefixes, to denote orientation:

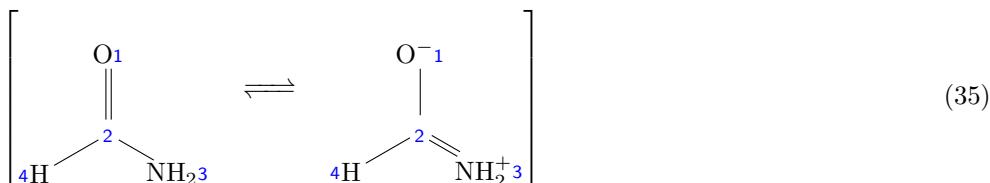
- **R vs. S** R is clockwise around a stereocenter, S is counterclockwise. This convention is used for chiral centers.

- **E vs. Z** E is high priority groups on opposite sides of the double bond. Z is high priority groups are on the same side of the double bond. This convention is used for double bonds.
- α vs β means that the hydroxyl group attached to C₁ (atom 8 in Figure 34 for both molecules) and the –CH₂OH group at C₅ (the groups on atoms 2 and 10 in Figure 34 for both molecules) lies on opposite sides of the ring's plane (a *trans* arrangement), while β means that they are on the same side of the plane (a *cis* arrangement). This convention is used when discussing the glycosidic bond between sugar molecules.



The molecule on the left is lactose with a β glycosidic bond, while the molecule on the right is maltose with an α glycosidic bond.

The nitrogen atom in amides have planar geometry even if it doesn't appear like it! This is because of the resonance with the carbonyl oxygen group, as shown here:



In this case, we can see that due to resonance, all of the atoms are coplanar for the amide. Quite separately, consider the following example problem:

20. The STRONGEST hydrogen bond is formed between:

- A the lone pair of O and a hydrogen bonded to O.
- B the lone pair of N and a hydrogen bonded to O.
- C the lone pair of O and a hydrogen bonded to N.
- D the lone pair of N and a hydrogen bonded to N.

The answer to this question is answer choice B. This is because the strongest hydrogen bond is formed when the hydrogen is extremely electron deficient (i.e. through being bonded to an oxygen) and it is hydrogen-bonded to something that is extremely basic (or in other words, is very good at donating its electrons in order to create the bond, which best describes the lone pair on a nitrogen atom).

3.2 Isomers and Stereochemistry

Isomers can be constitutional/structural (meaning they differ in connectivity of bonds) or stereoisomers (meaning they differ in the spatial arrangement of the atoms). Stereoisomers can either be conformers or configurational isomers. Conformers can differ by orientation in space, and are identical after a specific rotation about a σ -bond. Configurational isomers also

differ by orientation in space, but you can't rotate them to become identical. Configurational isomers can be categorized as either enantiomers or diastereomers. **Enantiomers** are nonsuperimposable mirror images, while **diastereomers** are nonsuperimposable structures that are not mirror images. Configurational isomers can also be categorized as either optical isomers or geometrical isomers. **Optical isomers** rotate plane-polarized light and cannot be rotated to become identical due to the asymmetry in the structure, while **geometrical isomers** are structures with limited rotation, and can't be rotated to become identical due to the presence of a ring or π -bond. Geometrical isomers are sometimes also called *cis/trans* isomers. The two categorizations are not mutually exclusive.

The Chan-Ingold-Prelog Rules are used to determine the **absolute configuration** (*R* vs. *S*) for a stereocenter. The rules are as follows:

1. First, you must prioritize the substituents that are attached to the carbon of the stereocenter according to the atomic mass of the atom directly bonded to the chiral carbon (from heaviest atom to lightest atom).
2. Next, orient the molecule in such a way that the substituent with the lowest priority points behind the plane of the molecule.
3. Finally, draw a semicircular arc from substituent 1 through substituent 2 and on to substituent 3. If the arc is clockwise, then the stereocenter is referred to as *R*. If the arc is counterclockwise, then the stereocenter is referred to as *S*.

Optical rotation is a characteristic feature of enantiomers that are described by absolute configurations. If, say, the *R*-enantiomer of a compound rotates the light in a positive direction (meaning clockwise) by X degrees, then the *S*-enantiomer of the compound rotates light by X degrees in the negative direction. However, in actuality, the *R*-enantiomer doesn't always rotate light in the positive direction—it really depends on the identity of the compound of interest.

For compounds with multiple chiral centers, here is another way of thinking about enantiomers vs. diastereomers:

1. **Enantiomers** are configurational isomers in which *all* of the chiral centers in each molecule are different from one another.
2. **Diastereomers** are configurational isomers in which *at least one, but not all* of the chiral centers in each molecule is different from one another.

For *geometrical isomers*, which have different spatial arrangements about a π -bond, the prefix of E is given for trans orientation of the two highest priority groups, while Z is given for cis orientation of the two highest priority groups.

Meso compounds are individual structures which contain a mirror plane slicing through the middle of the compound and an *even number* of chiral centers symmetrically displaced about the mirror plane. The net optical rotation of a meso compound is 0 degrees because the opposing chiral centers on each half of the molecule cancel one another out, leaving no net rotation of plane polarized light. A meso compound may be identified by either an inversion center in the middle of the molecule, or a mirror plane through the middle of the molecule.

When a molecule contains more than one chiral center, the maximum number of stereoisomers increases exponentially with each new chiral center according to the equation

$$\text{Number of stereoisomers} = 2^n \quad (36)$$

where n is the number of chiral carbons in the molecule. There are less than 2^n stereoisomers if one of the possible structures is meso. If n is odd, however, there can't be a meso compound, meaning that there must be exactly 2^n stereoisomers.

Rotational about σ bonds allows for different conformers. The two extreme overall structures are known as *staggered* (where substituents on the first atom do not block the substituents on the back atom of the σ bond in question) and *eclipsed* (where substituents on the first atom do block the substituents on the back atom of the σ bond in question). Within staggered conformation, the terms *gauche* and *anti* describe the relative position of substituents on adjacent atoms. *Gauche* conformers have the two groups of interest (often the largest groups) having a dihedral angle of 60 degrees, while *anti* conformers have the two groups of interest having a dihedral angle of 180 degrees. Staggered and anti is the most stable, because steric repulsion is minimized. To understand these different types of conformers, take a look at the Newman projections shown in Figure 3.2.

Cycloalkanes are basically cyclic alkanes that have one ring with the chemical formula C_nH_{2n} and contain no π -bonds. Three- and four- membered rings are reactive, while five- and six- membered rings are stable. The reactivity of tree- and four- membered rings is attributed to *ring strain*, defined as the energy difference between the linear and cyclic alkanes of equal carbon length. Let's look at a couple of particularly important cycloalkanes:

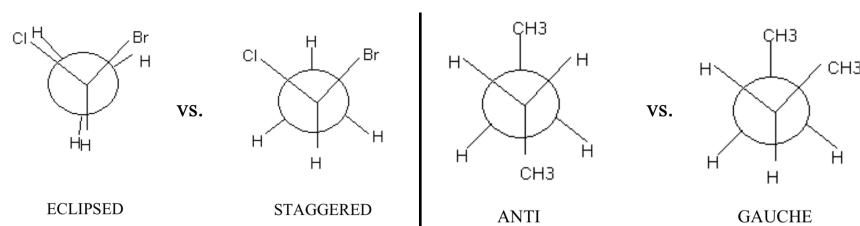


Figure 26: *Left* Staggered vs. eclipsed conformers. *Right* Staggered conformers can be either gauche or anti.

- Cyclopentane does not require much distortion of its bonds and shape to accommodate the 109.5° angle for the sp^3 hybrid. To achieve the correct angle and alleviate the torsional strain, cyclopentane forms an *envelope shape* where one of the carbons is not coplanar with the other four. However, even with this envelope shape, the substituents on the ring that are in an eclipsed conformation as a result of the near-planar ring structure are unfavorable, causing further contortion of the structure, which accounts for the ring strain energy. The envelope conformation is shown in Fig. 3.2.
- Cyclohexane has the most stable ring structure of all of the cycloalkanes. The most stable form is the *chair* conformation. Cyclohexane can flip between different chair conformations in a process called *ring-flipping*, and the interconversion process requires that it passes through the *boat* conformation. Equatorial positions (in the plane of the molecule) are more stable than axial (above and below the plane of the molecule), so the most stable conformation of a cyclohexane compound has the largest substituents in the equatorial positions. The chair conformation of cyclohexane is shown in Fig. 3.2.

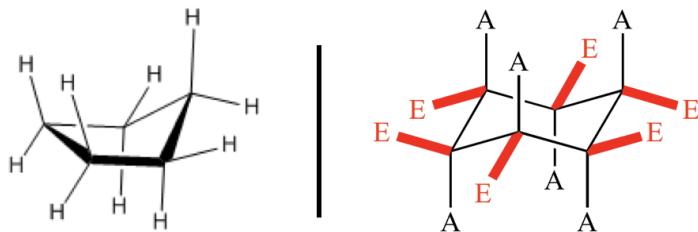


Figure 27: *Left* Envelope shape of cyclopentane. *Right* Axial (A) and equatorial (E) positions in cyclohexane chair conformation.

Separating Stereoisomers: In organic chemistry, there are compounds known as *chiral auxiliaries*, which introduce chirality to, or exaggerate existing chirality within, a reactant molecule. Chiral auxiliaries serve in a similar fashion to an enzyme. When aiming for one specific stereoisomer, it is easiest to select for it in the reaction. If not, then a mixture of stereoisomers forms and chirally selective separation techniques must be used. Chirally selective separation techniques come in two types: using an enzyme/chirally selective molecule to react specifically with one stereoisomer within the mixture, or invoking chirality in an existing separation technique. As an example for the first technique, if a new functional group is added to only stereoisomer by an enzyme, the two enantiomers now have different physical properties and can easily be separated. Once separated, the same enzyme can be employed to return the compound back to its original form. As an example for the second technique, a column chromatography gel can be made from a pure stereoisomer. If the column is made with an *R*-alcohol for instance, then when a racemic mixture of alcohols is added, the *S*-enantiomer has a greater affinity for the column, and thus has a greater elution time.

3.3 Nucleophilic Substitution

The strength of a leaving group can be predicted by the pK_a of its conjugate acid. The more stable the leaving group, the less basic the leaving group, and thus the more acidic the conjugate acid of the leaving group. In other words, the strength of a leaving group increases as the pK_a of its conjugate acid decreases.

A *racemic mixture* is a product mixture that has an even distribution of enantiomers. It is the observed product when

the mechanism involves an intermediate where the reactive site is an sp^2 -hybridized carbon and the molecule is symmetric (i.e. has no other chiral centers).

S_N2 reactions involve the nucleophile attacking prior to the leaving group leaving in a *backside* attack, essentially pushing the leaving group off of the electrophile. It has a trigonal bipyramidal transition state. S_N1 reactions involve the leaving

| Reactant Features | Course of Reaction Features | Product Features |
|---|--|---|
| The reactivity preference in an S_N2 mechanism is $1^\circ > 2^\circ > 3^\circ$ in terms of electrophiles. An S_N2 mechanism is favored with a good nucleophile. | An S_N2 mechanism forms a five-ligand transition state during the middle of the reaction. The 5-ligand transition state is the highest energy state and it exists for just a split second. It <i>cannot</i> be isolated. Steric forces destabilize the transition state by forcing bond angles to values less than 109.5° . | A single enantiomeric product is formed (i.e. no racemic mixture). S_N2 reactions exhibit second order kinetics. S_N2 reactions are one-step reactions, so they have fast rates of formation. |
| An S_N2 mechanism is favored in polar, aprotic solvents such as ethers and ketones. | | |

group leaving before the nucleophile attacks. The carbocation intermediate has a long enough lifetime to be detected using spectroscopy. Both rearrangement (e.g. hydride shifts and alkyl shifts) and a mixture of stereoisomers (formed from either front-side or back-side attack of the sp^2 -intermediate) are observed with S_N1 reactions. Formation of the carbocation is the rate-limiting step.

The S_N1 reaction can be complicated by **rearrangement** because of the carbocation intermediate formed. If a secondary

| Reactant Features | Course of Reaction Features | Product Features |
|--|---|--|
| The reactivity preference in an S_N1 mechanism is $3^\circ > 2^\circ > 1^\circ$ in terms of electrophiles. An S_N1 mechanism is seen with a poor nucleophile. | Steric hindrance pushes the leaving group off of the electrophile. | A racemic mixture forms when the electrophile has chirality. |
| An S_N1 mechanism is favored in protic solvent such as alcohols. | The intermediate is a planar, three-ligand carbocation where the carbon has sp^2 -hybridization. An intermediate is observed in addition to transition states. | S_N1 reactions exhibit first order kinetics. S_N1 reactions are slow, two-step reactions. |

carbocation (R_2CH^+) is formed, it can rearrange to form a more stable tertiary carbocation (R_2C^+) when possible. See Figure 3.3 for an example. Furthermore, if the electrophile has a chiral center at a site other than the electrophilic carbon, an S_N1 reaction will form both a major and minor product. The major product results from the transition state with least steric hindrance. Primary electrophiles proceed by S_N2 while tertiary electrophiles proceed by S_N1 . Secondary electrophiles

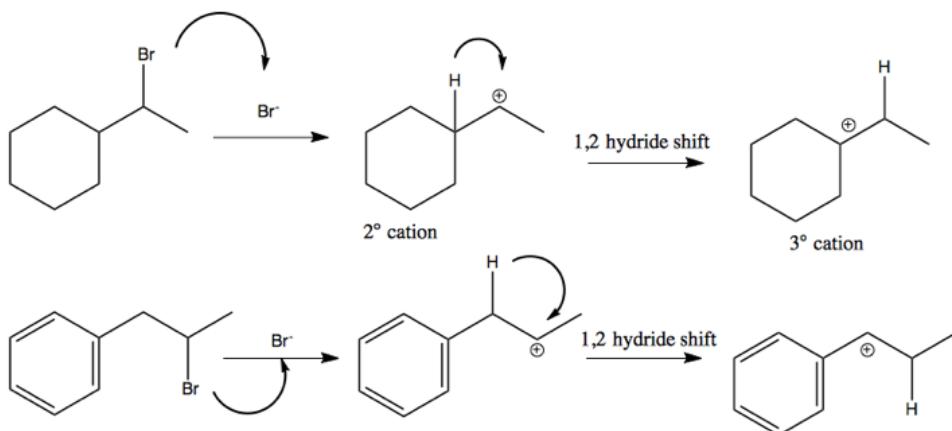


Figure 28: Example of favorable hydride shift to generate tertiary carbocation from secondary carbocation intermediate.

can proceed by either S_N 1 or S_N 2. Also, the stronger the nucleophile, the more likely the reaction will proceed by an S_N 2 mechanism; the better the leaving group, the more likely the reaction will proceed by an S_N 1 mechanism. Also, consider the following rate laws:

$$\begin{aligned} \text{S}_N 1 \text{ Rate} &= k [\text{Electrophile}] \\ \text{S}_N 2 \text{ Rate} &= k [\text{Nucleophile}] [\text{Electrophile}] \end{aligned} \quad (37)$$

Lastly, you can also consider the solvent to distinguish between S_N 1 and S_N 2 reactions: if the solvent is protic (i.e. capable of forming hydrogen bonds), the reaction will have a tendency to proceed by an S_N 1 mechanism. If the solvent is aprotic (i.e. *not* capable of forming hydrogen bonds), the reaction will have a tendency to proceed by an S_N 2 mechanism.

3.4 Structure Elucidation

When deducing the molecular structure for an organic molecule, it helps to know something about the symmetry of the compound and its units of unsaturation. The equation for units of saturation:

$$\text{Units of Unsaturation} = \frac{2(\#C) + (\#N) - (\#H) - (\#X) + 2}{2} \quad (38)$$

In terms of spectroscopy, the MCAT includes infrared (IR) absorption spectroscopy, ultraviolet-visible spectroscopy, ¹H NMR spectroscopy, and mass spectroscopy. We will discuss all of these.

Infrared Spectroscopy: Infrared spectroscopy starts by adding a monochromatic beam of IR photons to either a thin oil suspension (if the compound is a solid) or a neat solution (if the compound is a liquid) between salt plates. The molecule absorbs electromagnetic radiation that causes transitions between *vibrational energy* levels within it. Vibrational potential energy is described by the equation PE = $\frac{1}{2}kx^2$; we can say the spring constant *k* describes the bond strength, and the displacement *x* describes the distance from equilibrium that the bond has stretched. This means stronger bonds absorb at higher wavenumbers; this can be accomplished by either having multiple bonds or by bonding to an sp² carbon instead of an sp³ carbon, among other ways. IR spectroscopy is used as a diagnostic tool to detect certain functional groups.

Hydrogen bonding results in a broadened peak, observed in both the IR and NMR techniques. In the case of IR, the broadening of the –OH absorbance associated with hydrogen-bonding is caused by the weakening of the covalent bond between the hydrogen and the atom (nitrogen, oxygen, or fluorine) to which it is bonded. This lowers the energy of the covalent bond and thus lowers the energy of absorption for the bond. As the hydrogen bond increases in strength, the covalent bond weakens.

Ultraviolet-Visible Spectroscopy: While infrared photons affect the vibrational energies of a molecule, ultraviolet and visible photons affect the electronic energy levels. Because σ-bonds are so much stronger than π-bonds, the lowest energy absorbance for alkanes is significantly higher than the lowest energy absorbance for alkenes. As a result, we typically use UV-visible spectroscopy to analyze molecules with π-bonds, especially conjugated systems. UV-visible spectroscopy focuses on transitions between the π and π* energy levels. For systems with conjugation, there are several π-levels, but we care about only the lowest energy transition.

The transition of interest is from π to π*. The wavelength of highest absorbance, known as λ_{max}, changes with the amount of conjugation. As the amount of conjugation increases, so does λ_{max}. This is depicted in Fig. 3.4. Color results from excessive conjugation within a molecule. Conjugated aldehydes and ketones have about the same π-π* absorbances as conjugated alkenes of the same number of π-bonds. However, conjugated aldehydes and ketones have other, more intense absorbances that are of longer wavelength than their hydrocarbon counterparts. This is attributed to the n-to-π* transition associated with aldehydes and ketones, possibly because of the lone pair of electrons on the carbonyl oxygen.

Unlike IR spectroscopy, UV-visible spectroscopy can be applied in a quantitative fashion, and can be used to determine the yield of a reaction, if it involves a UV-visible active compound.

Nuclear Magnetic Resonance: Normally, all of the nuclei have spins of the same energy. However, when an external magnetic field is applied, spins can either align with the field or align against the field, so multiple energy states are possible. In the case of ¹H, there are two energy levels: α (the one aligned with the external magnetic field) and β (the one antialigned with the external magnetic field). The β energy level is defined as higher than the α energy levels. The energies of the two levels depend on the strength of the external magnetic field and the magnetogyric ratio of a particular nucleus.

Any nucleus with an odd number of protons or an odd number of nucleons has a net spin. ‘Spin’ is just that as the nucleus precesses, it generates a weak magnetic field. In the absence of any surrounding electrons, all identical nuclei exhibit the same spin and therefore require the same energy for excitation in an external magnetic field. Within a molecule, however, two

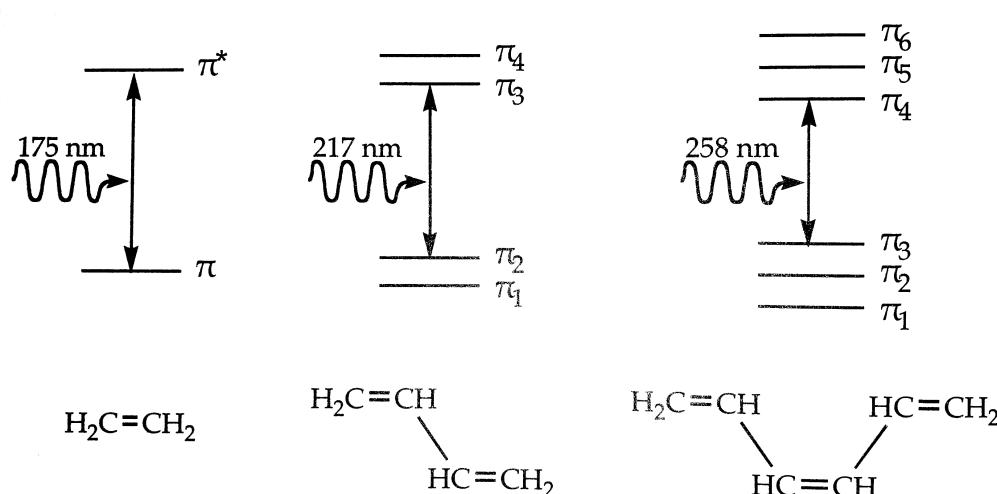


Figure 29: Comparing the wavelengths of highest absorbance for different molecules of varying degrees of conjugation.

identical nuclei may be in different electronic environments. As a result of the difference in their local magnetic fields, caused by the moving electrons, they do not require exactly the same amount of energy to excite the nucleus to a higher-energy spin state. ^1H NMR shift values are listed relative to a standard compound, tetramethylsilane ($(\text{H}_3\text{C})_4\text{Si}$). All twelve hydrogen atoms on tetramethylsilane are equivalent, so they absorb at the same value.

The best place to begin NMR for the MCAT is with molecular symmetry. Based on the symmetry of a molecule, you can determine the number of equivalent hydrogens that it contains. The number of different hydrogens in the molecule corresponds to the number of different peaks on the NMR graph. The number of hydrogens that are the same is also correlated with the peak intensity/height on the NMR graph.

In the case of alcohols, such as butanol, the protic hydrogen can be distinguished from other signals by its broadness. Broadening results from hydrogen-bonding in solution. Hence, alcohols are easily distinguished from ethers by the presence of a broad peak in their ^1H NMR spectrum.

The three basic components of an NMR graph are the integral, splitting pattern, and shift value. The *integral* is determined by the number of hydrogens making up a signal. The *splitting pattern* is derived from the coupling between hydrogen on adjacent neighboring atoms. The splitting pattern is also referred to as *coupling*. The *shift value* is determined by the local magnetic field caused by either lone pair electrons in motion or the electronic density associated with electronegative atoms.

Let's talk a bit more about splitting. Like electrons, nuclear particles have spin that can be classified as either up or down. The magnetic field resulting from the nuclear spin of hydrogen can be felt by the hydrogens on a neighboring atom. Because the spin can be either of two ways, the magnetic field may be additive or subtractive. The number of splitting peaks is equal to the number of equivalent hydrogens on the adjacent atoms plus one. The ratio of intensities of the peaks can be determined through binomial expansion (e.g. a quartet has a ratio of 1:3:3:1).

The shift value is used to assess the local electronic environment around a hydrogen. It is measured in parts per million (ppm) relative to the magnetic field necessary to detect H's on a standard compound, tetramethylsilane. Shift values tell us what functional groups are present. Deshielding effects require a stronger external magnetic field to energize the spin levels. The more deshielded a proton's environment is, the more *downfield* it is with a greater ppm.

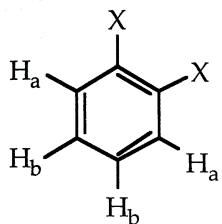
Be aware of certain peaks and features that occur over and over. For instance, whenever you see a triplet and quartet in a 3:2 ratio, you should conclude that there is an isolated ethyl group (H_3CCH_2^-) in the molecule somewhere. Whenever you see a doublet and a septet in a 6:1 ratio, you should conclude that there is an isolated isopropyl group ($\text{H}_3\text{C}_2\text{CH}^-$) somewhere in the molecule.

Distinguishing Disubstituted Benzenes: Integrals tell us the number of equivalent hydrogens in a signal and are often employed to determine the position of substituents on disubstituted benzene rings. Structures that are highly symmetrical have more equivalent hydrogens than asymmetrical structures. Fig. 3.4 shows the expected patterns on disubstituted benzene rings depending on if they are ortho-, meta-, or para- substituted. Para substitution is the easiest arrangement to distinguish of the three possible structural isomers, because it has a doublet of doublets. Para coupling is a highly recognizable feature.

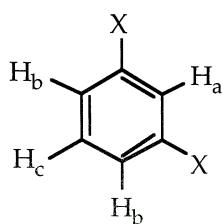
| Hydrogen Atom | δ (ppm) | Hydrogen Atom | δ (ppm) |
|--|----------------|---------------------------|----------------|
| RCH ₃ | 0.8-1.0 | RCH = CR ₂ | 5.2-6.4 |
| RCH ₂ R | 1.3-1.8 | RNH ₂ | 1-3 |
| RCOCH ₃ (ketone) | 2.1-2.5 | RNHCH ₃ | 2.0-3.2 |
| RC≡CH | 2.5-2.6 | ROH (alcohol) | 1-5 (broad) |
| ROCH ₃ (ether) | 3.5-4.0 | ArH (benzene) | 7.0-7.4 |
| RCH ₂ X (X=Cl, Br, I) | 3.0-3.8 | RCOH (aldehyde) | 9.0-9.8 |
| RCO ₂ CH ₃ (ester) | 3.5-4.0 | RCO ₂ H (acid) | 10-12 (broad) |

Deuterated Solvents for ^1H NMR: Because the solvent in ^1H NMR is in substantially higher concentration than

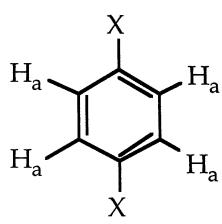
Case 1: The two substituents on benzene are equal:



2 different Hs
in a 1 : 1 ratio

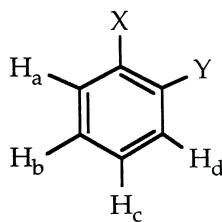


3 different Hs
in a 1 : 2 : 1 ratio

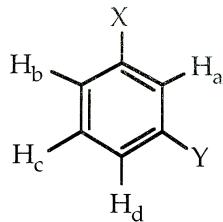


All Hs are
equivalent

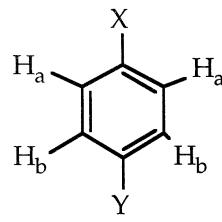
Case 2: The two substituents on benzene are not equal:



4 different Hs in
a 1 : 1 : 1 : 1 ratio



4 different Hs in
a 1 : 1 : 1 : 1 ratio



2 different Hs
in a 1 : 1 ratio

Figure 30: Discussing the expected ^1H NMR peaks for different benzene isomers.

the solute, it is imperative that the solvent not have any hydrogens. If the solvent has ^1H nuclei, then it would produce the largest signal in the spectrum. To avoid this problem, solvents are chosen that have deuterium ^2H instead. However, one potential problem with this occurs when protic compounds are dissolved into deuterated protic solvents, such as D₂O. Protic hydrogens can undergo exchange with the protons of the solvent, if the solvent is also protic. Although the dissociation constant may be small for compounds such as alcohols, over enough time all of the hydrogens can be released and then are able to reform their bonds. If D₂O is present in the solution, then deuterium will gradually replace protic hydrogens capable of undergoing exchange. This causes the signal for the protic hydrogen to disappear gradually.

Mass Spectrometry: In mass spectrometry, first the molecule is first converted into a cation following the addition of some input energy. The cation is then accelerated through an electric field after which it is deflected along the circular path by a perpendicular magnetic field. The radius of the circular path correlates to the particle's mass-to-charge ratio. In general, when ionizing a hydrocarbon, a bonding electron is lost from a C–H bond. Because cations and free radicals are unstable, the species rearranges and fragments, sheering off pieces of the molecule.

The base peak is the most intense peak, and the parent peak is the peak of the original molecule without any fragmentation. In mass spectroscopy, the material being analyzed should be at low pressure (so that the reactive free radical and cationic particles do not collide and undergo reactions), and also in the gas phase (so that the particles travel independently of one another in the apparatus).

Mass spectrometry is particularly useful in detecting chlorine and bromine, because for both of these there are two predominant isotopes in a high concentration. This ratio is easily recognized when viewing the results from mass spectrometry. Elemental bromine is about 50% ^{79}Br and 50% ^{81}Br , so mass spectrometer results for compounds with bromine show two peaks of roughly equal size separated by 2 atomic units (amu). This is a telltale feature of a compound that contains a single bromine atom. Elemental chlorine is about 75% ^{35}Cl and 25% ^{37}Cl , so mass spectrometer results for compounds with a single chlorine show two peaks with a 3:1 size ratio separated by 2 atomic mass units. No other elements common in organic molecules have isotopes separated by two amu with such large population percentages.

3.5 Lab Techniques

Lab techniques can be broken down into three categories: separation, purification, and identification.

Separation of molecules requires that at least one of the molecules be in motion. We shall use the word *flow* to describe the motion of bulk material. The ability to flow depends on the state of matter. When two materials are in different phases, they can flow in different directions and therefore can be separated. A common sense view of separation techniques dictates that if you have a mixture of two or more components, you must convert the system into a state where the components are in different phases.

Distillation removes a liquid from either another liquid or from solute by exploiting their boiling point differences. Upon heating, the most volatile component converts to a gas more readily than the less volatile components. This vapor distillation purification cycle is repeated multiple times until the desired purity is achieved. This is what is referred to as *fractional distillation*. If it is only done once, it is called *simple distillation*. Simple distillation is employed with liquids that have a large difference in boiling points, and also to remove a solvent from a solute. Simple distillation is faster than fractional distillation and it generates a higher yield, while fractional distillation leads to greater distillate purity. A typical distillation apparatus is shown in **Fig. 3.5**.

Chromatography is the separation of two or more components in a mixture by exploiting their differences in (1) solubility in a migrating solvent and (2) their affinity for a polymer. The mobile phase is one of several possible solvents, ranging from nonpolar to polar/protic, or in some instances a mixture of two solvents. The stationary phase is one of two materials: alumina or silica gel. Both of these polymers are polar, so in the absence of any other information, we know that polar materials have a higher affinity for the stationary phase than nonpolar materials. As a consequence, polar species have slower migration rates than nonpolar species in chromatography.

Thin Layer Chromatography (TLC) is carried out on a small scale to identify the number of components or type of components in a mixture. It involves spotting small aliquots of sample near the bottom of a vertical, rectangular plate with either silica gel or alumina on its surface in a line parallel to the base of the plate. This plate is then placed into a container that can be fitted with a lid and solvent is added to the container until it is just below the level of the spots. Due to capillary action, the solvent slowly migrates up the plate, interacting with each of the spots as it travels. In order to prevent the solvent from evaporating, a lid is employed to maintain a closed system. Once the plate is developed, one can ascertain where the spots ended, and thus know how many components were in the mixture and whether the components were nonpolar, semi-polar, or highly polar. A typical TLC set up is shown in **Fig. 3.5**.

In TLC, we often measure an R_f value for the compounds, which we can think of as a ‘ratio of the fronts,’ where the

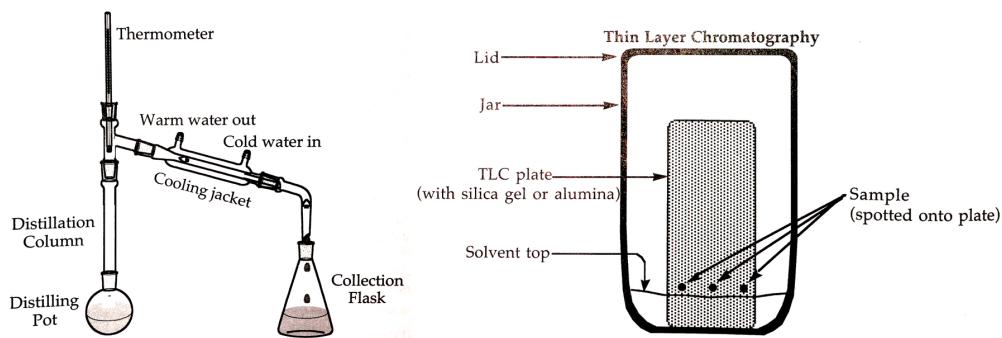


Figure 31: *Left* A typical distillation apparatus/setup. *Right* A typical thin layer chromatography setup.

distance a spot travels is compared to the distance that the solvent travels. The component with the greatest R_f value is

the component that has the least affinity for the stationary phase (adsorbent) and is the most soluble in that solvent. The range for R_f values is $0 \leq R_f < 1$, because there is a chance the spot does not travel at all or that it travels almost at the solvent front. Because a solute dissolves best into a solvent of ‘like’ nature, the higher the value of R_f value, the more that the solute is like that solvent. Based on R_f values, you should be able to make predictions as to the best solvent for column chromatography. The best separation occurs when the R_f values have the greatest ratio. The basic idea of all forms of chromatography is *separation by relative affinity*.

Column chromatography is applied to separate bulk quantities of product. It operates on the same principles as TLC: the solubility of a compound in the solvent as it flows down the column versus the affinity of the compound for the adsorbent in the column. If the compound migrates quickly down the column, we can conclude that it must experience a minimum attraction to the silica gel or alumina (stationary phase) in the column. The sample and solvent flow down the column; faster migrating compounds reach the bottom of the column first, so they have shorter elution times. A large R_f value corresponds to a low elution time, given that the compound has a fast migration rate in each case.

Chiral columns have become popular in recent years, because they are capable of selecting for one enantiomer over another. The most common designer polymers on the MCAT to date are the adsorbents used in ion exchange chromatography. Three with which you should be familiar are **sulfonated-polystyrene**, **carboxymethyl-cellulose (CM-cellulose)**, and **diethylaminoethoxy-cellulose (DEAE-cellulose)**. The first two carry negative charges at neutral pH, while DEAE-cellulose carries a positive charge at neutral pH. These allow us to isolate the oppositely charged proteins based on electrostatic interactions.

Gas chromatography (GC) works by first vaporizing a sample (mixture of components) into the gas phase and then forcing that organic vapor through a packed column using elevated pressure from a cylinder of an inert gas. The machine measure retention time on the column by recording collisions as gas molecules leave the end of the long, coiled column and strike the detector. The collisions are converted to peaks, which can be integrated to determine their relative abundance. The compounds in the sample must be highly volatile at 200°C and inert with regards to silica gel or was for GC analysis to work. **Fig. ??** shows the design for a standard gas chromatography machine. Given that the compounds vaporize instantaneously upon addition to a GC machine, the boiling point does not affect the elution time. However, because many columns are often slightly polar, affinity for polar species is greater than affinity for nonpolar species. A second factor that dictates the migration rate is mass—heavier gases move slower and have longer elution times. heavier gases and polar gases typically have high boiling points, so although boiling point has little to do with elution time, the factors that raise boiling point also increase the elution time in a GC.

Extraction works based on solubility and it’s typically employed to take advantage of drastic difference in the solubilities of components in two different (immiscible) solvents. First, we need to define **partitioning**. A compound has a different solubility in every solvent. The ratio of its maximum solubility in one solvent compared to its maximum solubility in another solvent is known as the *partition coefficient*. In cases where the compound can go into either solvent, it will select between the solvents in a ratio equal to the partition coefficient. When the solvent system involves two immiscible liquids, then the solute must split between the two solvents. This means that the components of a mixture can be separated if they have different partition coefficients. By decanting one solution from the other, the components are then separated from one another in their solute form according to their relative solubilities in the two different solvents.

When one of the two immiscible solvents is water, the solubility properties of the solutes can vary with the water pH. **Acid-base extraction** separates compounds by their pK_a values. Here is an example problem:

To separate benzoic acid ($\text{C}_6\text{H}_5\text{CO}_2\text{H}$) from toluene ($\text{C}_6\text{H}_5\text{CH}_3$), it would be best to use a mixture of an:

- A alcohol and water at $\text{pH}=11$.
- B alcohol and water at $\text{pH}=3$.
- C ether and water at $\text{pH}=11$.
- D ether and water at $\text{pH}=3$.

To carry out an extraction, the two solvents cannot be miscible in one another. Alcohols are typically miscible in water, so choices A and B are eliminated. Benzoic acid can be deprotonated in basic solution, where it becomes its anionic conjugate base (benzoate), which is readily soluble in water. Toluene is nonpolar and thus dissolves into the organic ether rather than water, so the best answer is choice C.

On the MCAT, perhaps the most difficult task when it comes to extraction will be analyzing a flow chart. **Fig. 3.5**

shows a typical four-compound flow chart. It is a good general rule that a weak base should be added before a strong base, so as to isolate the stronger acid. If you were to add the sodium hydroxide first, it would extract both the phenoxide and the carboxylate into the aqueous layer. Also, because extraction involves both acidic water and basic water, there is the potential for hydrolysis. As such, extraction cannot be used to isolate water-sensitive compounds such as acid anhydrides, acid halides, and esters.

Purification: Many of the separation techniques we have discussed so far can also be used as purification techniques.

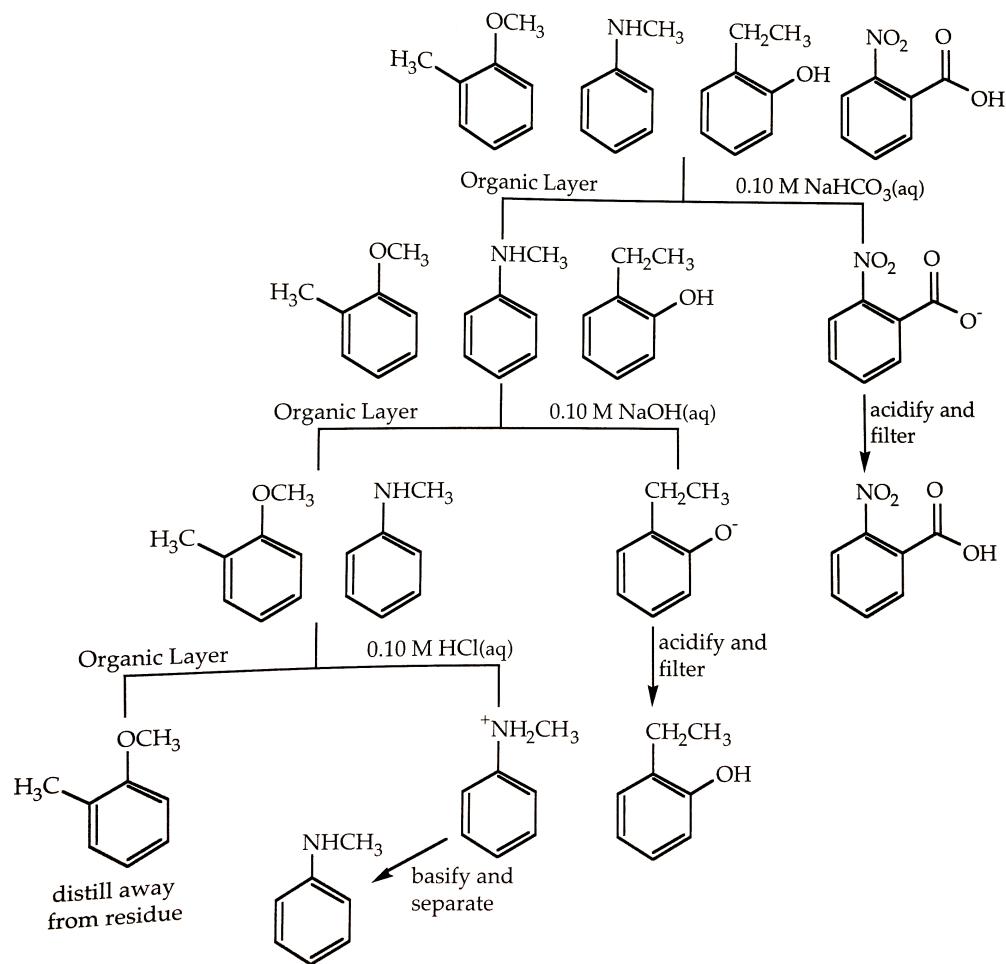


Figure 32: An example extraction flow chart.

For example, column chromatography is useful for purification, because the bands separate nicely from one another. A solid product may be separated from solid impurities through selective precipitation, which is referred to as **recrystallization**. Recrystallization involves dissolving the solid into hot solvent, hot filtering out the insoluble solid impurities, and then slowly cooling the solution to precipitate the purified crystals. Upon cooling the solution, the product precipitates out of solution in a crystalline form free of the impurities from before. The crystals are filtered at low temperature, to prevent any of the material from dissolving back into solution, to avoid any reduction in the yield.

Solvent choice is essential in recrystallization. In order to maximize the amount of crystal collected, the desired material should be insoluble (or minimally soluble) in the solvent at lower temperatures. Sometimes the solvent can be a homogeneous mixture of two solvents to meet desired solubility properties.

In some instances, the purification can be enhanced by adding a material to preferentially bind some or all of the impurities. One such case occurs when you have colored impurities and a colorless target crystal. By adding activated charcoal (charcoal that is free of impurities), the colored species will be preferentially bound by the surface of the charcoal and can thus be filtered out. Charcoal binds all organic materials, but because of its π -network and the presence of conjugated π -networks in colored species, charcoal has a higher affinity for colored organic molecules than colorless organic molecules.

Filtration is employed to remove solids from liquids. A fine microscopic net is used to separate the solid from the liquid by allowing the liquid to flow through while hindering the flow of solids, which are larger than the pore size of the filter. Filtering typically depends on gravity to force the liquid to flow through the filter, although it can be expedited by a vacuum pressure difference.

The final step of crystallization is cold filtering (cold to prevent the loss of product to dissolving) and washing the precipitate with a solvent that does not dissolve the target compound, but dissolves the impurities. It is optimal to use a solvent with a low boiling point, so it can readily evaporate away from the solid. Here are a couple practice problems regarding purification:

If a solid is too soluble in water at room temperature, it would be best to add which of the following solvents?

- A Hexane
- B Ethanol
- C Diethyl ether
- D Tetrahydrofuran

The key to this question is that the solvent you mix with water must reduce the overall polarity and hydrogen bonding capacity of the solvent mixture *and* be soluble in water. All of the choices are less polar and less protic than water, but of the choices, only ethanol is soluble in water. Choice **B** is therefore the best answer.

Which of the following mixtures can be separated using gravity filtration?

- A Ethanol from diethyl ether.
- B Sodium chloride from water.
- C Butane from dichloromethane.
- D Oxalic acid from hexane.

In choice A, both compounds are liquids, so filtering is pointless. In choice B, the salt is completely soluble in water, so the salt is actually in solute form. It must be a solid, not solute, to be filtered, so choice B is eliminated. Butane is a liquid as is dichloromethane, eliminating choice C, leaving choice **D** as the correct answer. Hexane is a liquid and all carboxylic acid, besides formic acid and acetic acid, are solids at room temperature. Acids do not readily dissolve into hydrocarbon solvents, so oxalic acid exists as a solid and not as a solute in hexane.

Identification Techniques: You can (somewhat) identify/verify a compound or ascertain its purity from its melting point. A rapid change from a solid into a liquid is indicative of a pure compound. Here is an example problem:

If a student suspects that the compound they have is L-tyrosine, what should they add to the sample for a mixed melting point?

- A L-Tyrosine
- B D-Tyrosine
- C Xylose
- D Glycine

To support the notion that an unknown compound is L-tyrosine using a mixed melting point study requires mixing the unknown with a pure sample of L-tyrosine. If the unknown compound is in fact L-tyrosine, then the mixture is in fact pure L-tyrosine. This would result in a sharp melting point at the characteristic melting point of L-tyrosine. If the unknown compound were not L-tyrosine, the melting point range would be broad and would fall outside of the characteristic melting point of L-tyrosine.

Chemical tests involve reactions that are selective for very few functional groups. The ideal chemical test is positive

for only one functional group, but this is rare. Most chemical tests have multiple functional groups that give positive results. For a chemical test to be effective, a few things must be true:

1. The reagent must react with very few functional groups.
2. The chemical testing reagent must undergo some visible change, such as a color change or phase change, so it can be detected.
3. The testing reagent must be the limiting reagent, so that any excess of the testing reagent will not interfere with the reading.

Derivatives are products formed from the reaction of a given reagent with an unknown compound. Derivatives are made to be used as a diagnostic for an unknown. The additional physical measurements associated with the newly formed derivative can be used to identify both the derivative and indirectly the unknown. They are used to support and confirm an identity for an unknown. Common derivatives include 2,4-dinitrophenylhydrazones for ketones or aldehydes, benzylesters for alcohols, and osazones for sugars. The ideal derivative is a solid that can easily be identified by its melting point.

3.6 Lipids

Lipids are natural molecules rich in long-chain alkyl groups that share many properties with hydrocarbons, especially solubility. Lipids include fats, fat-soluble vitamins (in particular vitamins A, D, E, and K), some hormones, sterols (steroid alcohols), waxes (esters with a long-chain fatty acid combined with a long-chain alcohol), and fatty acid esters of glycerol such as monoglycerides, diglycerides, triglycerides, and phospholipids. There are insoluble in water and thus require assistance when being transported through the body. We shall define lipids as any bioorganic molecule that is not soluble in water.

Fatty acids are carboxylic acids that have a long, aliphatic chain. Most naturally occurring fatty acids contain an even number of carbons and are unbranched. This is because the natural building block for fatty acid biosynthesis is the 2-carbon species, **acetyl CoA**—its structure is shown as a part of the synthesis pathways shown in Fig. 3.6.

Waxes are the esters of fatty acids, involving long-chain alcohols. The result is a highly hydrophobic ester with a high melting point. **Soaps and detergents** are amphipathic molecules, which means they contain both a hydrophobic region and a hydrophilic region. As a result, they act like surfactants when added to water.

Fatty Acids Derivatives and especially straight-chain fatty acids are naturally synthesized from **acetyl-CoA** and **malonyl-CoA** via a five-step cycle that adds two carbons to the chain and then is repeated until the fatty acid backbone is complete. From there, the fatty acid has a few potential pathways, including being stored as a triglyceride or becoming part of a phospholipid. This is shown in Fig. 3.6.

From here, the fatty acid can take one of two pathways: (1) adding to glycerol-3-phosphate to become part of a cell

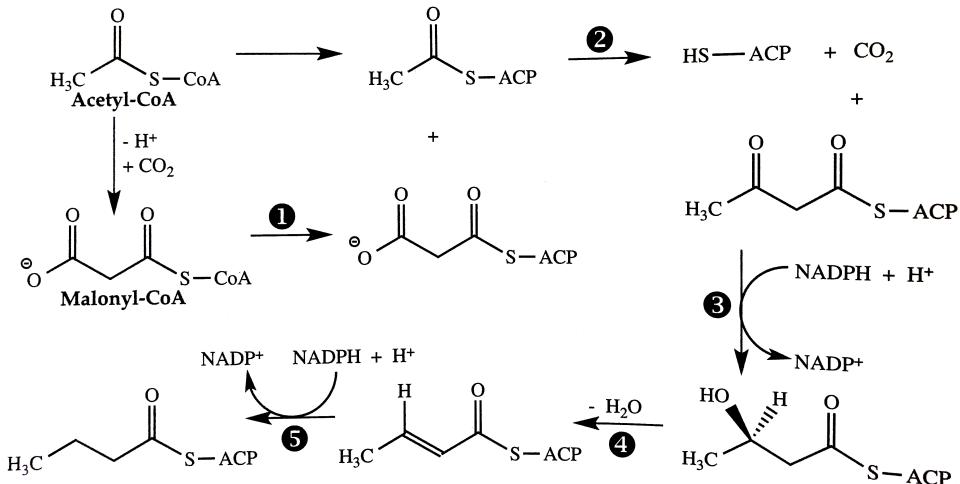


Figure 33: Five-step cycle for the synthesis of a fatty acid backbone.

membrane or (2) adding to a glycerol to become a triglyceride (stored fat in adipose tissue). Fig. 3.6 shows the synthesis

of a phospholipid, phosphatidate. The formation of fatty acid triglycerides is similar to the formation of a phospholipid, although the final product is completely hydrophobic. **Fig. 3.6** shows the complete synthesis of a fatty acid triglyceride from dihydroxyacetone phosphate.

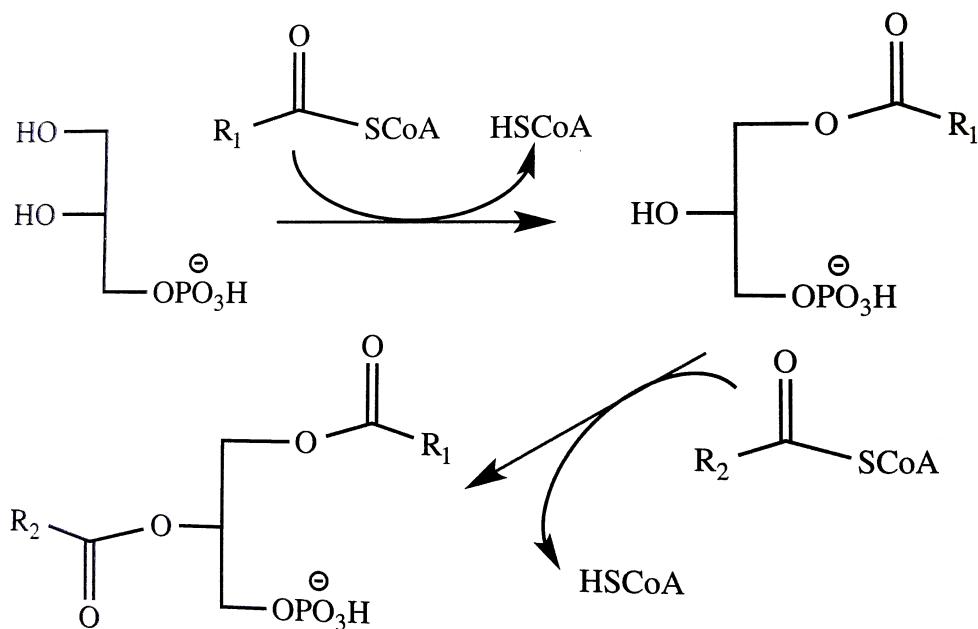


Figure 34: Synthesis of phosphatidate, a phospholipid. In general, the first R-group (R_1) is fully saturated while the second R-group (R_2) is often unsaturated. The exact fatty acids depend on the cell membrane for which the phospholipid is destined.

3.7 Lipid Signal Molecules

Eicosanoids are 20-carbon compounds derived from **arachidonate** that act as *local hormones*. Eicosanoids include **prostaglandins**, **thromboanes**, and **leukotrienes**.

Prostaglandins all share a five-membered ring from C8 through C12. They are given names PGA through PGI, where the letter refers to specific functional groups and stereochemistry. The numerical subscript refers to the number of π -bonds not in the ring. Eicosanoids are all derived from **arachidonate** through oxidation to common intermediates PGG₂ and PGH₂. Drugs such as aspirin act on the *cyclooxygenase* enzyme that is involved in the synthesis of PGG₂. Prostaglandins are shown in **Fig. 3.7**.

Steroids are a class of biological compounds that contain the *androstane ring* structure (the specific four-ring structure) with various alkyl groups and functional groups. They serve many biological functions, including emulsifying fats, regulating cell membrane fluidity, and as signal molecules for many body functions. **Sterols** are a subgroup of steroids that contain hydroxyl groups, which is common in several hormones.

Vitamins are nutrients that we must consume through diet. We classify vitamins according to their solubility as either *water-soluble* vitamins or *fat-soluble* vitamins. Water-soluble vitamins are readily passed through the body and excess quantities can be excreted as waste. Fat-soluble vitamins, on the other hand, are hard to clear from the body and therefore reach toxic levels if consumed to excess. Four fat-soluble vitamins are Vitamin A, Vitamin D, Vitamin E, and Vitamin K.

Terpenes and *terpenoids* (biological molecules derived from terpenes) are natural hydrocarbons found in plants and animals that are made from 5-carbon isoprene (2-methyl-1,3-butadiene) units. The five-carbon skeleton of isoprene can be found in terpenes. Terpenes are classified by their number of carbon atoms. Monoterpenes have 10 carbon atoms, sesquiterpenes have 15 carbons, diterpenes have 20 carbons, sesterterpenes have 25 carbons, and so on. Studies in biogenesis show that the large terpenes are synthesized using *isopentenyl pyrophosphate* rather than isoprene. Pyrophosphate adds across the diene of isoprene to form either isopentenyl pyrophosphate or dimethylallyl pyrophosphate, which are interconverted by isomerization. **Fig. 3.7** shows the chemical structures of isoprene, isopentenyl pyrophosphate, and dimethylallyl pyrophosphate.

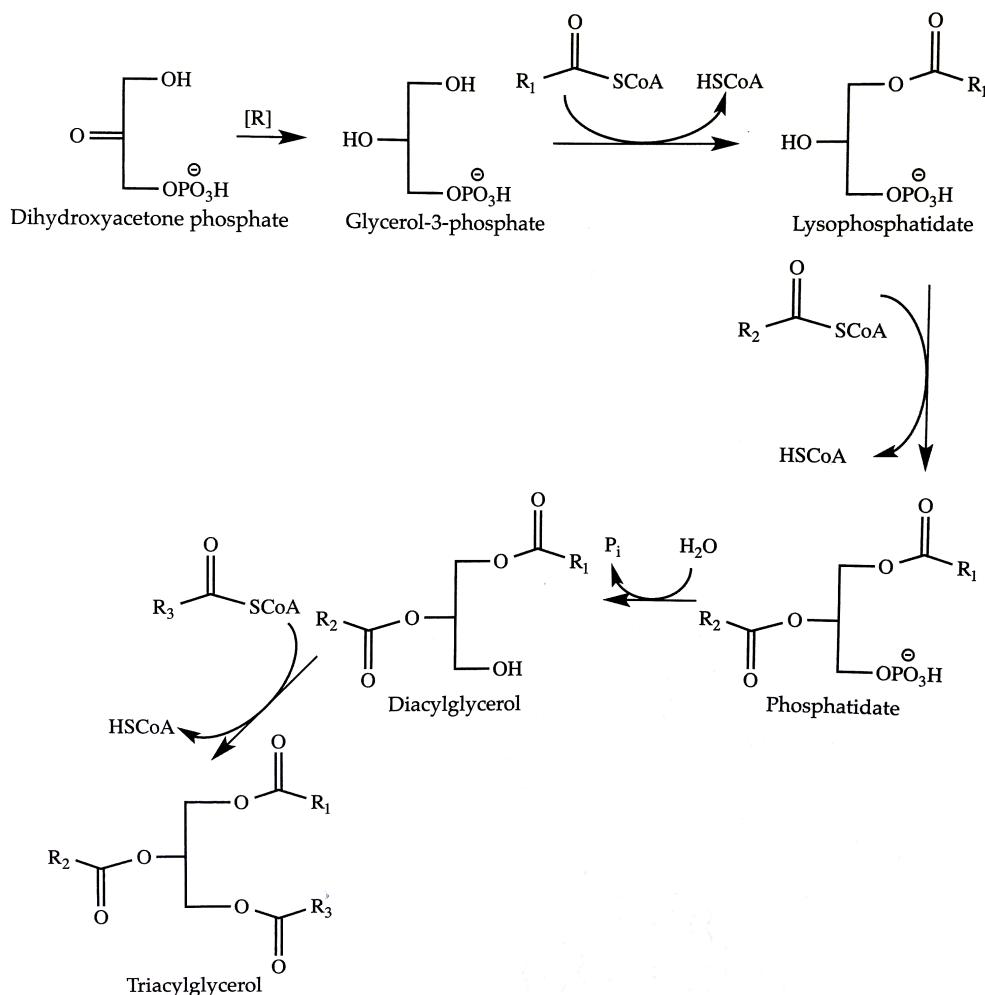
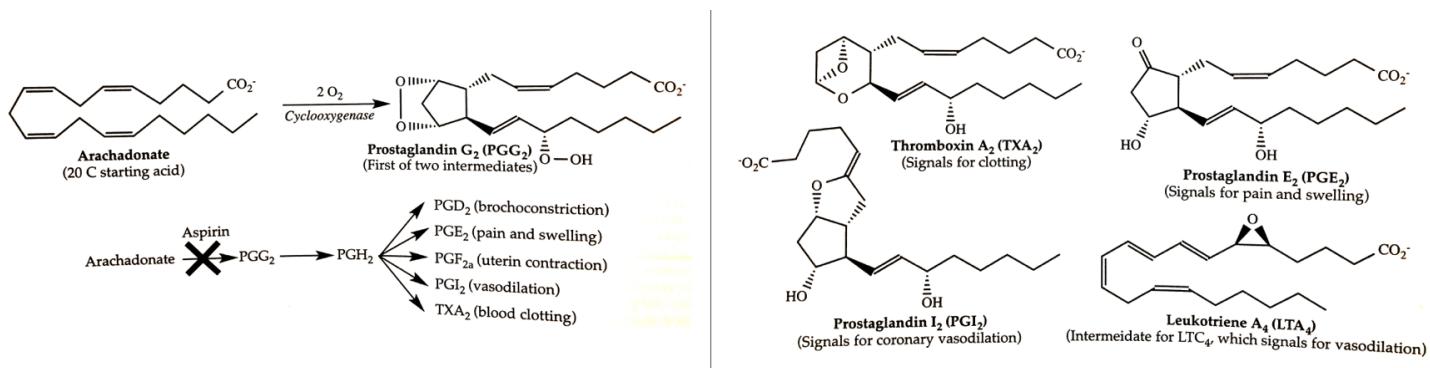


Figure 35: Synthesis of the fatty acid triglyceride from dihydroxyacetone phosphate.

Figure 36: *Left* Some sample eicosanoids. *Right* Synthesis of eicosanoids from arachidonate.

Both plants and animals synthesize terpenes.

The synthesis of cholesterol involves terpenes and terpenoids, even though cholesterol itself is not a terpene. As a general rule, smaller terpenes are found primarily in plants, while some larger terpenes like lanosterol and β -carotene are found in both plants and animals. One of the skills that you must develop to do well on terpene-related questions on the MCAT is to

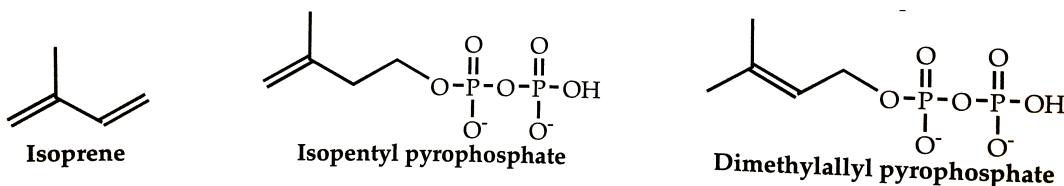


Figure 37: The chemical structures of isoprene, isopentenyl pyrophosphate, and dimethylallyl pyrophosphate.

recognize terpenes and be able to identify the isoprene subunits in the carbon skeleton. [Fig. 3.7](#) shows the analysis of some terpenes for the isoprene units in the skeletal fragments.

Terpenes are UV-active, because of their π -bonds and conjugated system. As the degree of conjugation increases, both

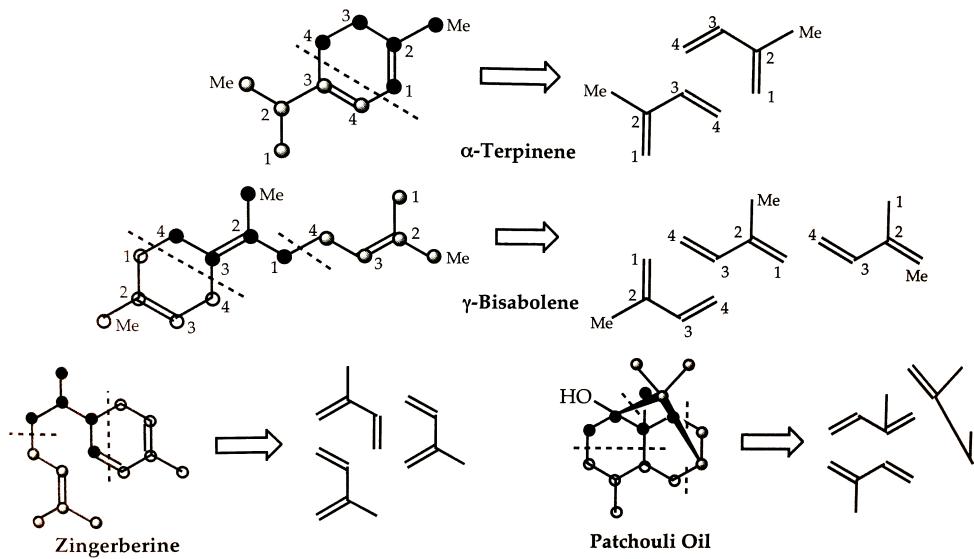


Figure 38: Identifying terpenes for their isoprene units in the skeletal fragments.

the intensity and λ_{max} increase. A common lab technique employed to isolate terpenes is *steam distillation*, where the pulp of a natural material is placed into water and boiled so that the natural oils are distilled from the pulp. This allows the essential oils to vaporize at a temperature lower than their boiling point, so they do not degrade. The distillate is a mixture of water and terpenes, which are easily separated using extraction techniques. Terpenes can also be extracted from pulp. When isolating terpenes, it is often a mixture of geometrical isomers that is collected. The different isomers of a terpene are given the prefixes α -, β -, γ -, and so on. The different geometrical isomers have similar physical properties, but because of differences in conjugation, they exhibit differences in the absorption of UV-visible photons.

Terpenes are often isolated in educational laboratory experiments. Because of their biological significance and the fact they are isolated in lab experiments, they are represented on the MCAT. If you have a fundamental understanding of terpenes, then you should be fine. Here is an example problem involving terpenes:

Which of the following compounds is NOT a terpene or terpenoid?
See [Fig. 3.7](#) for the answer choices.

Terpenes are built from 5-carbon subunits, which means that they should contain a carbon total that is a multiple of five. Choice D has sixteen carbons, so it cannot be a terpene or terpenoid.

3.8 General Bioorganic Reactions

There's no way you can memorize all of the reactions that can be tested on the MCAT, so we choose to present classes of reactions you should know, and in doing so will present some specific examples that hopefully seem familiar. You should

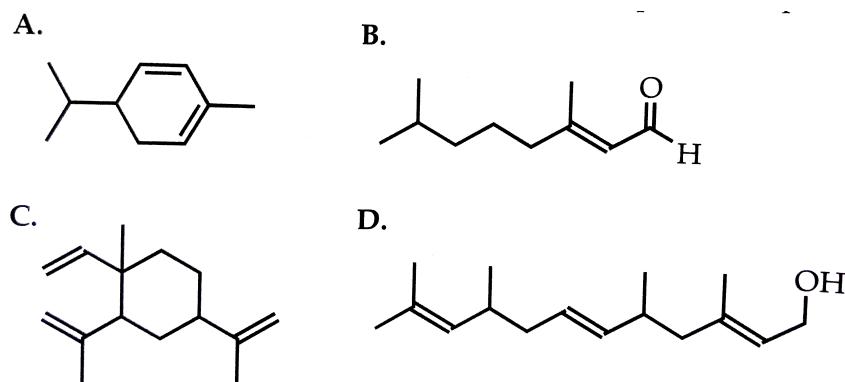


Figure 39: Identifying terpenes for their isoprene units in the skeletal fragments.

recognize (1) substitution reactions where a nucleophile displaces a leaving group from carbon (covered previously in Nucleophilic Substitution section of Organic Chemistry), (2) addition reactions where two groups (an electrophile and nucleophile) add to the two neighboring carbons of a π -bonds, (3) electrocyclic reactions where π -bonds and rings are formed or broken in a concerted fashion, (4) redox reactions where carbons gain and/or lose bonds to hydrogen and oxygen, and finally (5) free-radical reactions where carbon is electron deficient and thus highly reactive. The examples in this section will have biological application to show that the approach can also help with biochemistry as well as organic chemistry.

Reactions with alkenes are common in biochemistry. You should be familiar with hydrogenation, dehydrogenation, hydration, and dehydration. **Fig. 3.8** shows two simple biochemical reactions of alkenes.

Pericyclic reactions involve the repositioning of σ -bonds and π -bonds through a cyclic transition state. These reactions

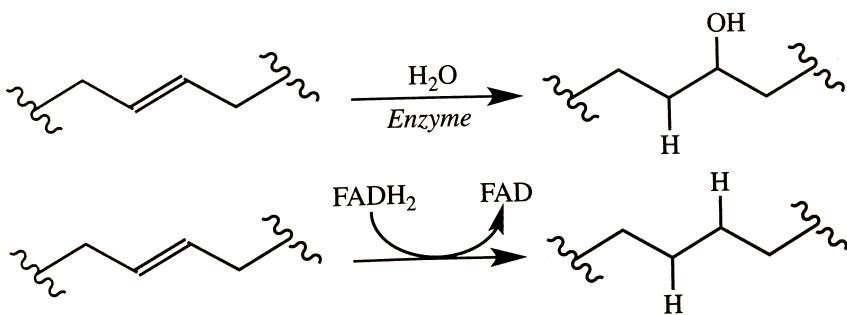


Figure 40: Biochemical examples of alkene addition reactions.

are believed to be concerted, meaning that the formation and breaking of all bonds occur simultaneously. Pericyclic reactions include cycloaddition reactions, sigmatropic rearrangement, and electrocyclic reactions. In cycloaddition, two separate compounds come together, resulting in a single new compound. In sigmatropic rearrangement reactions, it is an intramolecular rearrangement that takes place. An electrocyclic addition reaction involves the addition of a conjugated diene (with 4 π -electrons) to an alkene (with 2 π -electrons) to form a six-membered ring. The transition state is similar to the resonance of benzene, as shown in **Fig. 3.8**. These operations can be both ring closing and ring opening.

Biochemical redox reactions are also a common reaction type in biological systems. Examples include dehydrogenation, reactions involving NADH, NADPH, or FAD, and reactions involving methyl transfers.

3.9 Carbonyls and Alcohols

Fig. 3.9 shows several types of carbonyl compounds and carbonyl derivatives. More generally, oxygen-containing compounds tested on the MCAT include carbonyls, carbonyl derivatives, and alcohols. Because of their ability to form hydrogen bonds, alcohols typically have high boiling points and are generally miscible in water. Alcohols are not good nucleophiles, but they can be deprotonated and converted into an *alkoxide* (the deprotonated form of an alcohol) under basic conditions. Because

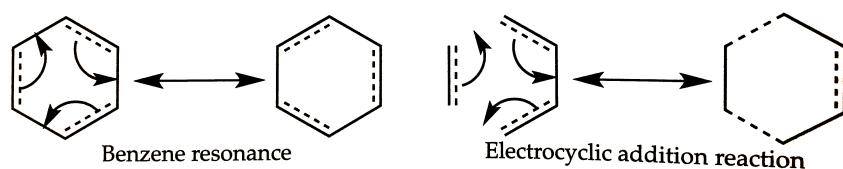


Figure 41: An example of an electrocyclic addition reaction. Notice its similarity to the resonance of benzene.

alkoxides are strong bases, they are not an ideal nucleophile, but they are generally better than alcohols. Alcohol chemistry also involves oxidation of the hydroxyl group into a carbonyl group. Alcohols can also be formed from the reduction of carbonyls. On ^1H NMR, the $-\text{OH}$ group does not couple well, so we rarely consider splitting patterns for alcohols. The $-\text{OH}$ peak also slowly disappears with the addition of D_2O to the NMR tube.

Because aldehydes and ketones do not form hydrogen bonds, they typically have boiling points only *slightly* higher than

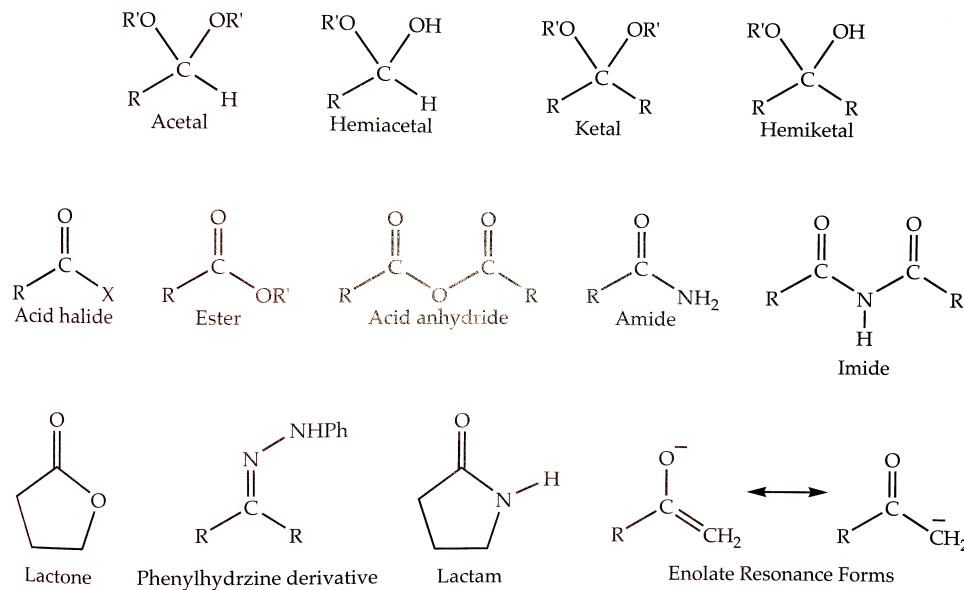


Figure 42: Selected carbonyl compounds and their derivatives.

alkanes of equal mass. Because of the polarity of the carbonyl bond, they are slightly miscible in water. The reactivity of aldehydes and ketones occurs primarily at the electrophilic carbonyl center. Aldehydes and ketones are reactive with most nucleophiles, but not by a ‘traditional’ nucleophilic substitution mechanism. Once a nucleophile attacks a carbonyl carbon, it forms a four-ligand intermediate with a negative charge on oxygen known as a *tetrahedral intermediate*. The chemistry of aldehydes is similar to the chemistry of ketones, except that an aldehyde can be oxidized into a carboxylic acid while ketones cannot be oxidized easily.

Ketals and **hemiketals** are derivatives of ketones while **acetals** and **hemiacetals** are derivatives of aldehydes. Ketals occur when a ketone loses the carbonyl group and gains two alkoxy $\text{R}-\text{O}$ functional groups. Notice that the oxidation state of carbon doesn’t change. A hemiketal occurs when the ketone has its carbonyl group converted into a hydroxyl group and gains one alkoxy group. Acetals are similar to ketals, except it is the aldehyde that loses its carbonyl group to gain the two alkoxy groups. Hemiacetals are similar to hemiketals, except it again is an aldehyde, rather than a ketone, that converts its carbonyl group into a hydroxyl group while gaining an alkoxy group. Fig. 3.9 shows the formation of these four compounds.

Ketals and acetals serve as protecting groups for carbonyl groups in synthesis involving ketones and aldehydes. Hemiacetals and hemiketals are not useful as protecting groups, but they are important in sugar chemistry. Acetals and ketals can be formed and removed only under acidic conditions, whereas hemiacetals and hemiketals are formed only under basic conditions but removed under any conditions. Fig. 3.9 shows the mechanism for formation of a ketal from a ketone under acidic conditions. It is the same mechanism for the formation of an acetal from an aldehyde, except that an aldehyde is the

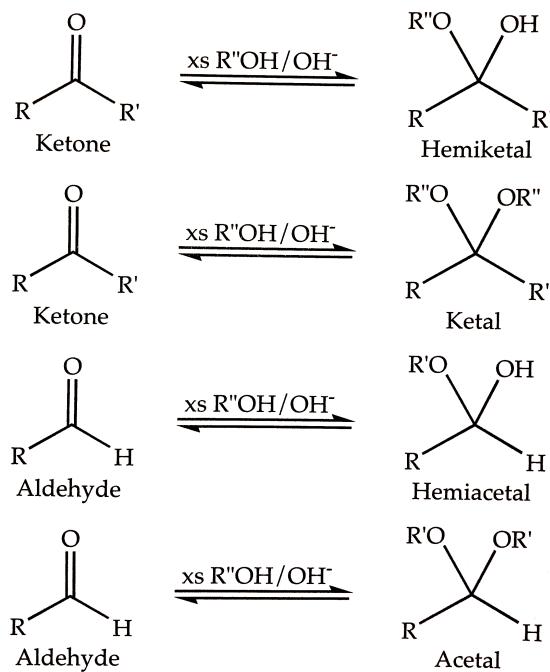


Figure 43: Formation of ketals, hemiketals, acetals, and hemiacetals.

reactant, rather than a ketone. Meanwhile, **Fig. 3.9** shows the formation of a hemiacetal from an aldehyde and alcohol in the presence of a strong base. These mechanisms should be kept as simplistic as possible.

Protecting groups are used in synthesis to prevent a reagent from reacting at a site where it is undesirable to have

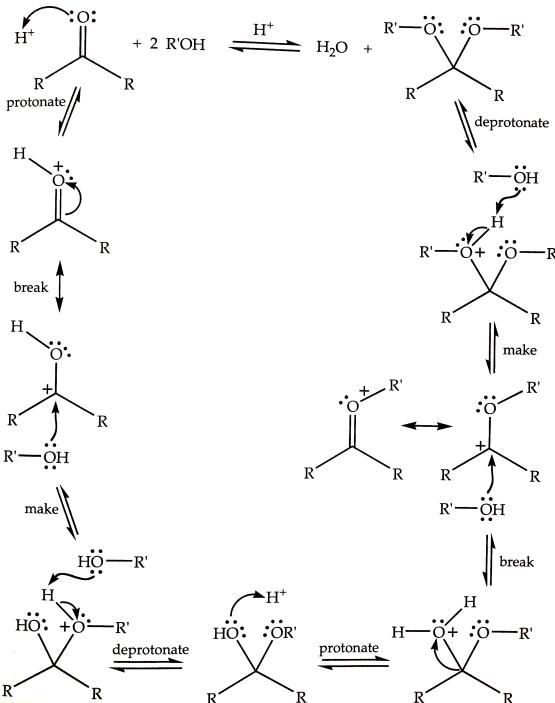


Figure 44: Mechanism for the formation of a ketal from a ketone under acidic conditions.

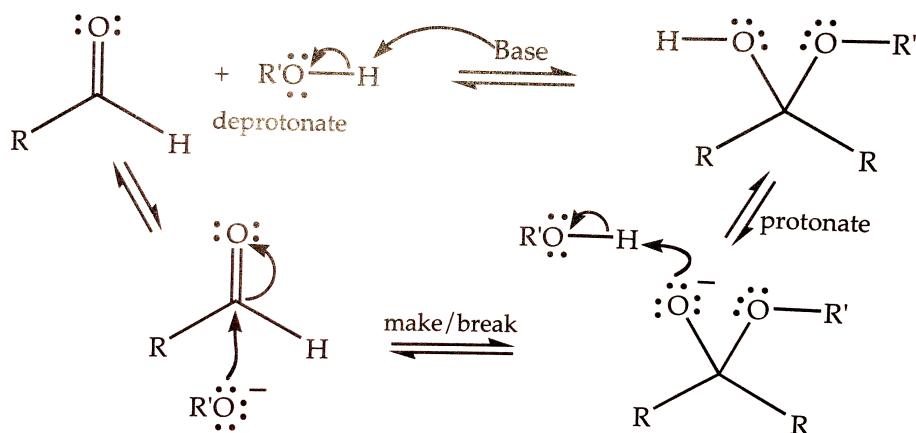


Figure 45: Mechanism for the formation of a hemiacetal from an aldehyde and alcohol in the presence of strong base.

a reaction. Aldehydes and ketones employ the same reaction to add the same protecting group, forming either an acetal or ketal. Because acetals and ketals are less reactive than aldehydes and ketones, they are an ideal protecting group. The alcohol that is typically used to form the protected carbonyl compound is a vicinal diol (a 1,2-diol), such as ethylene glycol ($\text{HOCH}_2\text{CH}_2\text{OH}$). Fig. 3.9 shows the protecting of cyclohexanone using ethylene glycol. As a guideline, any time that you have a molecule with more than one reactive site, you must protect the sites at which you wish to have no reaction. The exception to this rule is when the site you wish to react is significantly more reactive than any other sites on the molecule.

Unlike aldehydes and ketones, **carboxylic acid derivatives** have a functional group that potentially can act as a

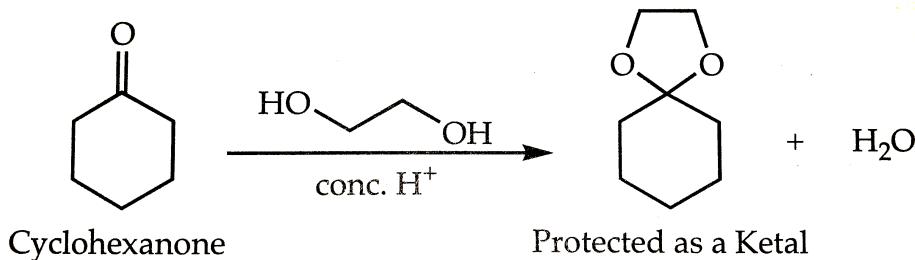


Figure 46: Ketals and acetals as protecting groups.

leaving group. Carboxylic acids can be converted into esters, anhydrides, or acid halides. Carboxylic acids can be formed by *saponification* (i.e. treating an ester with strong base in water), by treating a methyl ketone with I₂ and a strong base, by oxidizing primary alcohols and aldehydes in water, or by hydrolyzing a nitrile or an amide using strong acid at high temperatures. These reactions are shown in Fig. 3.9. Carboxylic acids can also be reduced into primary alcohols or converted into other compounds such as acid halides, acid anhydrides, or esters. Fig. 3.9 shows four reactions of carboxylic acids with which you are expected to be familiar with.

Esters are carbonyl compounds with an alkyl group and an alkoxy group attached to the carbonyl carbon. Esters have a leaving group (i.e. the alkoxy group), so they undergo more reactions than ketones or aldehydes. Esters can exchange their alkoxy group in the presence of acid and an alcohol in what is referred to as a *transesterification reaction*.

Lactones are cyclic esters. They can be synthesized by treating cyclic ketones with peroxyacids (RCO_3H) in what is known as the *Baeyer-Villiger* reaction. Fig. 3.9 shows the formation of a lactone from an intramolecular transesterification reaction of a hydroxyester.

Acid anhydrides are formed by a condensation reaction of two carboxylic acids at high temperature. Therefore, when you add water to an acid anhydride, it breaks into two carboxylic acids.

Acid halides are similar to esters, but with a halide (Cl, Br, or I) in place of the alkoxy group. They are the most reactive of all the carbonyl compounds, because the halide is a great leaving group. They undergo the same substitution reactions as other carbonyl compounds that have a leaving group, but they react faster.

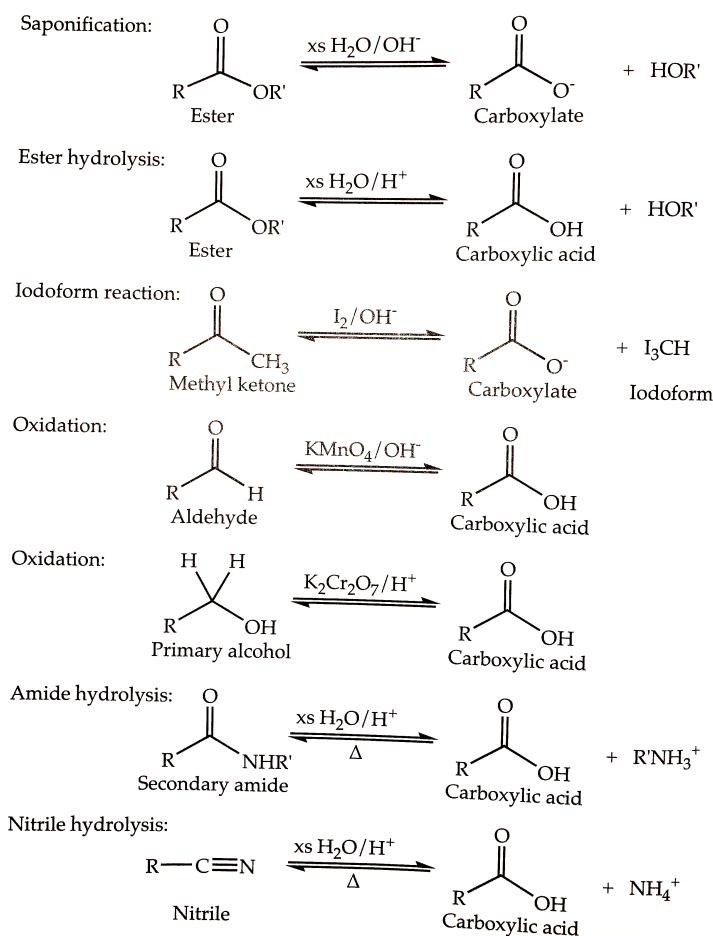


Figure 47: Different formation reactions for carboxylic acids.

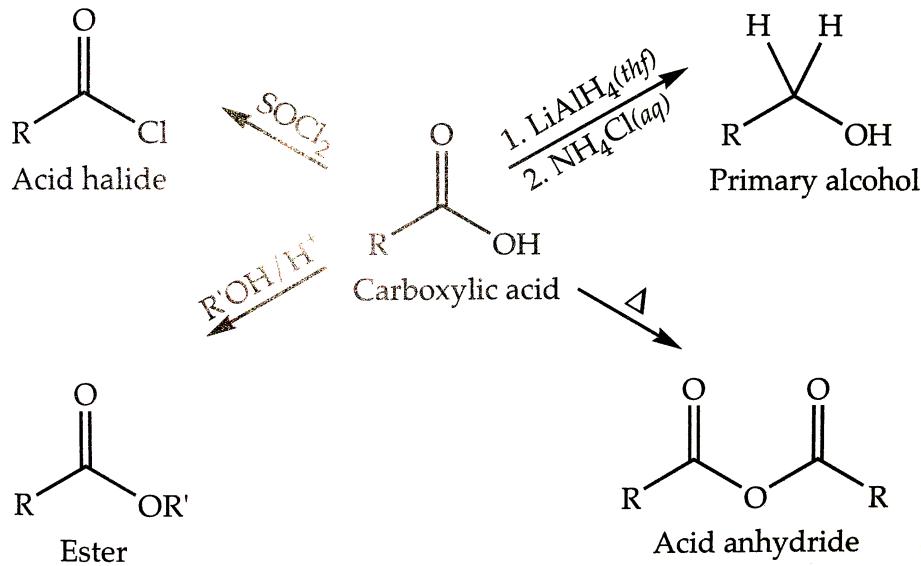


Figure 48: Four reactions of carboxylic acids with which you are expected to be familiar.

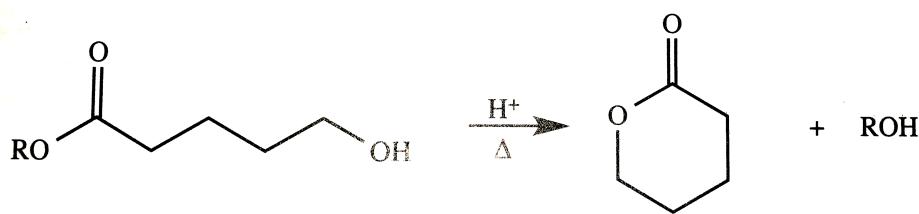


Figure 49: Formation of a lactone from an intramolecular transesterification reaction of a hydroxyester.

Amides are carbonyls with an amine group bonded to the carbonyl carbon. Amides form the backbone of proteins, and they are found in most of the bases of nucleic acids. An amide bond that links amino acids together is referred to as a *peptide bond*. Amides can be reduced into amines using a strong reducing agent, such as $LiAlH_4$.

3.10 Carbonyl Reactivity

Because carbonyl compounds contain a $C=O$ bond, they are good electrophiles. The difference in reactivity between a ketone and a carboxylic acid derivative, such as an ester, centers around the presence of a leaving group on the carbonyl carbon. A *tetrahedral intermediate* is formed when a carbonyl compound is attacked by a nucleophile. For carbonyl compounds that have leaving groups, the reactivity of the carbonyl compound is based on the strength of the leaving group—stronger leaving groups make for a more reactive (more electrophilic) carbonyl compound. For example, an acid halide can easily be converted into an anhydride, ester, or amide. An acid anhydride can readily be converted into an ester or amide, but it is hard to convert the anhydride into an acid halide.

The hydrogen on the α carbon (the carbon adjacent to the carbonyl carbon) is mildly acidic, so a strong base can remove it. **Enolates** form when a hydrogen on the α carbon is deprotonated. The enolate can gain a proton at the oxygen, forming an **enol**. A ketone and enol are **tautomers** (structural isomers that vary in the position of a π -bond and a hydrogen), so the conversion from a ketone into its enol is known as *tautomerization*, as shown in Fig. 3.10. The α carbanion that results when an H is stripped by a strong base is a good nucleophile, so when an alkyl halide is added to the solution, the carbanion can attack the alkyl halide in a nucleophilic substitution reaction to form a new carbon-carbon bond. This results in a longer ketone. The generic reaction is shown in Fig. 3.10.

Oxidizing agents cause oxidation (and in doing so, they get reduced). Oxidizing agents are rich in oxygen, poor in

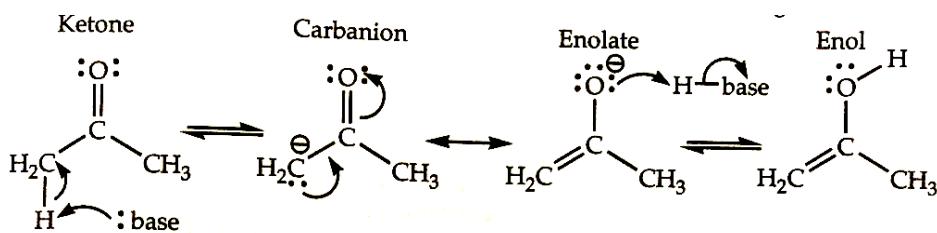


Figure 50: Keto-enol tautomerization.

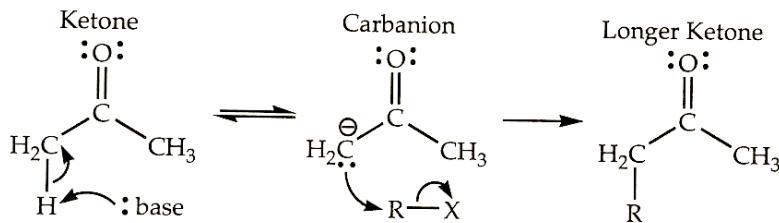


Figure 51: Carbanions in ketones to form carbon-carbon bonds.

hydrogen, and have an atom in a high oxidation state with high electron affinity. Examples include $KMnO_4$, RCO_3H , O_3 ,

NAD^+ , CrO_3 , ROOR , FAD , and NADP^+ . Reducing agents cause reduction (and in doing so, they get oxidized). Reducing agents are poor in oxygen, rich in hydrogen, and have an atom in a low oxidation state with low ionization energy. Examples include LiAlH_4 , HCl/Zn , H_2/Pd , NADH , NaBH_4 , $\text{H}_2\text{NNH}_2/\text{OH}^-$, FADH_2 , and NADPH .

Oxidation of a primary alcohol with mild oxidizing agent (e.g. PCC in pyridinium) gives an aldehyde. Oxidation of a primary alcohol with strong oxidizing agent (e.g. K_2CrO_7 followed by acid work up with H_2SO_4) gives a carboxylic acid. Oxidation of a secondary alcohol with strong oxidizing agent (e.g. KMnO_4 followed by base work up with KOH) gives a ketone. You can't oxidize a tertiary alcohol because it doesn't have any hydrogen's to lose. For other oxidation reactions, you should simply recognize the oxidizing agent and know that the product is more oxidized than the reactant. The common theme is that one of the reactants (i.e. the oxidizing agent) is rich in oxygen.

The two most common reducing agents in carbonyl chemistry are LiAlH_4 and NaBH_4 . LiAlH_4 is more reactive than NaBH_4 , and so NaBH_4 reduces only ketones and aldehydes, while LiAlH_4 reduces all carbonyl compounds. Two important reduction reactions are the aldehyde-to-alkane reactions: *Clemmensen reduction* and *Raney nickel reduction*, respectively. Both can also reduce a ketone to an alkane. These two reactions are shown in Fig. 3.10. Ketones can also be reduced to alkanes using H_2NNH_2 .

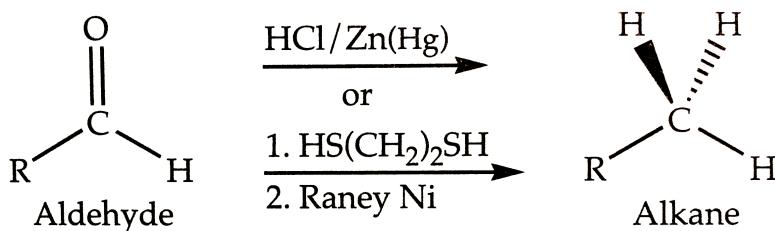


Figure 52: Clemmensen reduction (top arrow) and Raney nickel reduction (bottom arrow) for aldehyde reduction into an alkane.

3.11 Name Reactions and Synthesis Logic

There are specific name reactions that you should know, according to the AAMC's official guide.

Grignard reactions involve the addition of a carbanion in the form of a carbanion metal halide to the electrophilic carbon of a carbonyl to form a new carbon-carbon bond. Almost all Grignard reactions generate alcohols. No matter what the reactant, Grignard reactions generate a hydroxyl group in the product where the carbonyl group originated. Keep in mind that if the carbonyl has good leaving groups attached, then it is possible that the Grignard reaction can happen multiple times if there's an excess of the Grignard reagent. Remember that Grignard reagents are strong bases, so it is critical that no protic hydrogens are present. This is why the reaction uses anhydrous ether solvent. The second step in the Grignard reaction (i.e. the work-up step) requires a weak acid, as opposed to a strong acid, to avoid protonation of the hydroxyl group, which can lead to an E_1 elimination reaction. The Grignard reaction can also take place with CO_2 to form a carboxylic acid.

Aldol reactions start with the deprotonation of an α hydrogen from a carbonyl (ketone or aldehyde) using a strong base. The resulting anionic compound (enolate) nucleophilically attacks a neutral carbonyl compound (electrophile) to form a β hydroxy carbonyl compound. Under typical reaction conditions, it undergoes elimination to yield an α, β -unsaturated carbonyl. Regioselectivity is involved, because the final product can have either E or Z structural geometry. The anionic species can also be protonated on the oxygen, resulting in the formation of an enol, but that is in equilibrium with the carbonyl. The mechanism for aldol condensation is shown in Fig. 3.11.

When the ketone is asymmetric, the choice of which α proton to deprotonate is dictated by reaction conditions, namely steric hindrance of the transition state and thermodynamic stability of the intermediates and product. When the major product is selected to minimize steric hindrance in the transition state, this product is said to be the *kinetic product*. This term refers to the lower activation energy required to go through a less sterically hindered transition state. When the major product is selected to maximize stability of the intermediates or product, this product is said to be the *thermodynamic product*. The reaction has kinetic control at low temperatures with a strong, bulky base, such as t-butOK. The reaction has thermodynamic control at high temperatures with a strong, small base. Aldol reactions have a lot of biological relevance—for example, the enzyme responsible for the fourth step of glycolysis is an *aldolase*, which catalyzes the retro-aldol reaction. Let's take a look at the following example problem:

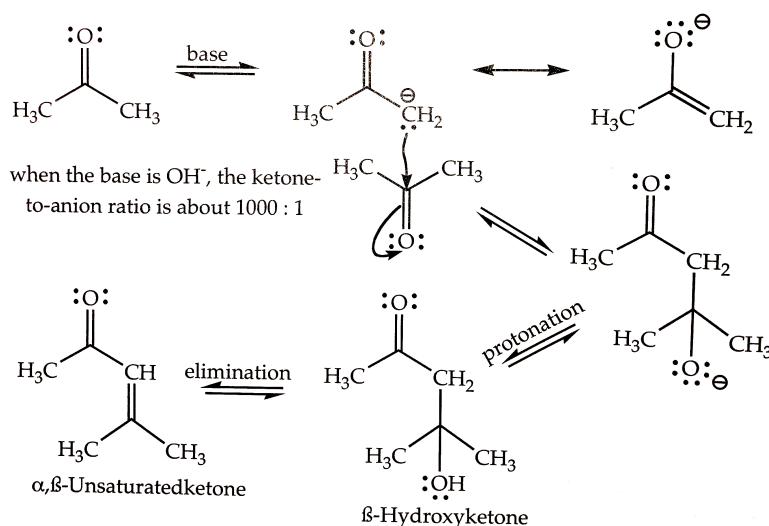


Figure 53: Aldol condensation mechanism.

The reaction shown in Fig. 3.11 is best described as:

- A a Claisen condensation via the kinetic enolate.
- B a Claisen condensation via the thermodynamic enolate.
- C an aldol condensation via the kinetic enolate.
- D an aldol condensation via the thermodynamic enolate.

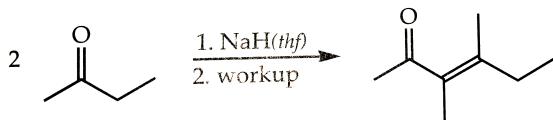


Figure 54: Figure for practice problem with aldol condensation.

The reaction involves the condensation of an asymmetric ketone upon itself, which eliminates choices A and B, because the Claisen condensation requires the reaction of esters with esters, not aldehydes or ketones. Because the more hindered α -carbon is involved in the addition, the reaction must have been carried out by way of the thermodynamic enolate (more substituted enolate). This makes choice **D** the best answer.

A **Claisen reaction** is just like the aldol condensation reaction, with the notable exception being that an ester is reacting rather than a ketone or aldehyde.

Transesterification involves the exchange of one alkoxy group for another on an ester. Because the equilibrium constant is roughly one, the reaction is driven by either the removal of a product or the addition of a reactant. Transesterification can be carried out in either acid or base. When acid is employed, the mechanism is a typical acid-catalyzed process. In base, the alkoxide nucleophile causes the reactivity. If the base is hydroxide, a carboxylate and alcohol are formed as the two products. The biological applications of transesterification include membrane transport and the conversion of fatty acids into triglycerides. Fatty acids are found in cell membranes; fatty acids that have been esterified with glycerol are broken down in basic water in a process known as *saponification*, generating glycerol and three fatty acid carboxylates.

The **Wittig reaction** converts a carbonyl into an alkene, with stereoselectivity. You can get either the E or Z product depending on what reagents you choose. The Wittig reaction is accomplished by a four-centered intermediate in its last step. See Fig. 3.11 for an example reaction and the associated critical part of the mechanism. The **Baeyer-Villiger** reaction converts a ketone into an ester using a peroxyacid. The reaction proceeds through an attack of the carbonyl carbon followed by an alkyl migration, which converts a ketone carbon into a more oxidized ester carbon. The **iodoform reaction** falls in

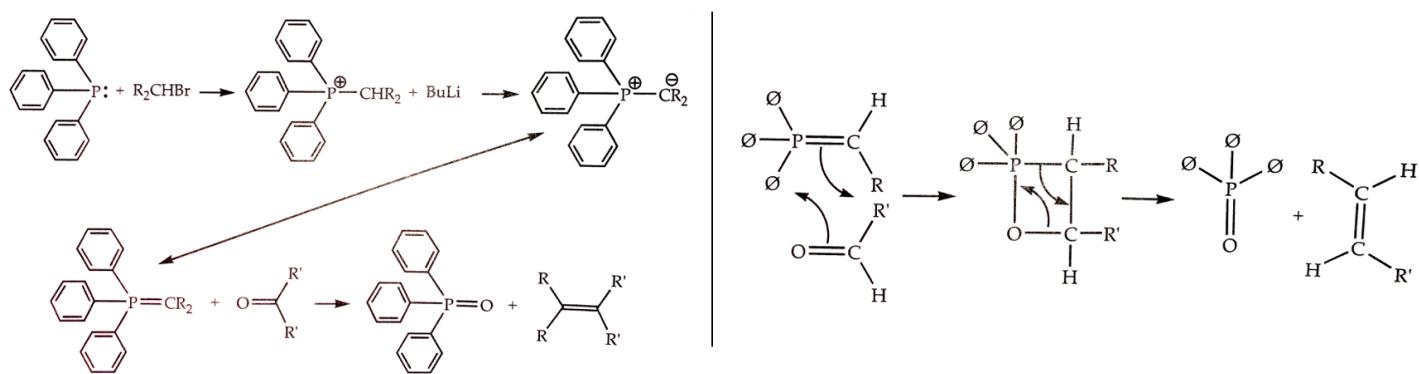


Figure 55: *Left* Example of the Wittig reaction for converting a carbonyl into an alkene with stereoselectivity. *Right* Four-centered intermediate in the last step. The \emptyset symbol is shorthand to represent a phenyl group.

a general class of reactions known as the haloform reactions, which convert a methyl ketone into a carboxylic acid and a haloform, HCX_3 . The iodoform reaction is used as a chemical test for a methyl ketone, because iodoform is yellow in color and insoluble in water. Methyl ketones react with iodine I_2 in the presence of strong base to generate a carboxylic acid and deprotonated iodoform molecule. The carboxylic acid is deprotonated to yield a carboxylate and iodoform. Examples of these two reactions are shown in Fig. 3.11.

The **Wölk-Kishner reduction** is a two-step reaction that converts a ketone or aldehyde into an alkane. It is carried

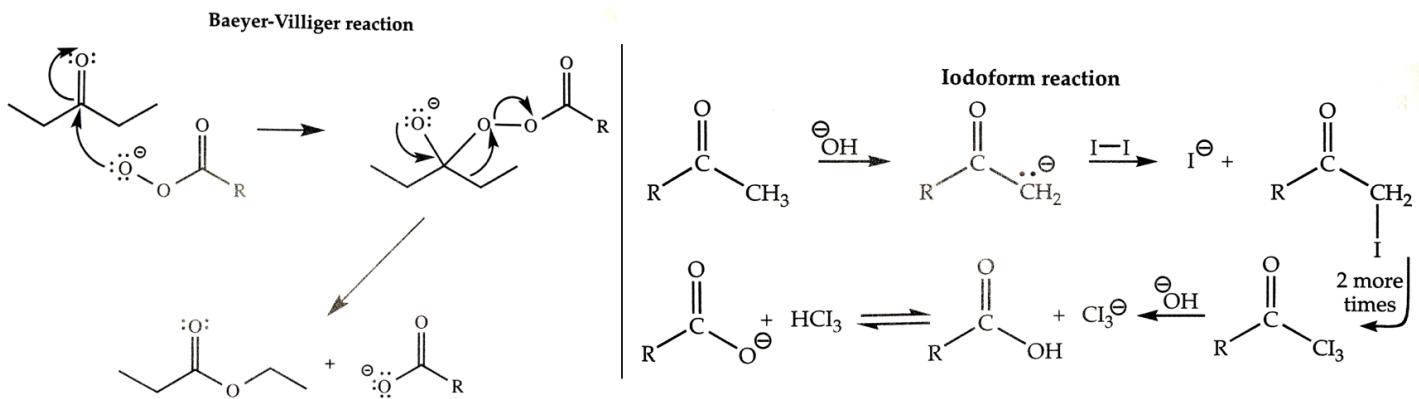


Figure 56: *Left* Example of the Baeyer-Villiger reaction for converting a ketone into an ester using a peroxyacid. *Right* Example of the iodoform reaction for converting a methyl ketone into a carboxylic acid.

out under basic conditions using hydrazine H_2NNH_2 followed by aqueous hydroxide. The general reaction is shown in Fig. 3.11. Step 1 of the Wölk-Kishner reaction substitutes a hydrazine for the carbonyl oxygen, resulting in the formation of a hydrazone and water by-product. The first step is reversible with the addition of acidic water, so you can regenerate the carbonyl by adding water and a catalytic amount of acid to the hydrazone. Carbonyls can also react with amines to generate an imine, as we shall see in the nitrogen compound section. Treating the hydrazone with strong base in water generates nitrogen gas and the alkane.

The steps to analyze a synthetic pathway are:

1. Identify the new bonds and functional groups formed by counting the atoms (particularly the carbons) and noting the changes in bonds from reactant to product.
2. Break the product into fragments that match the reactant skeleton.

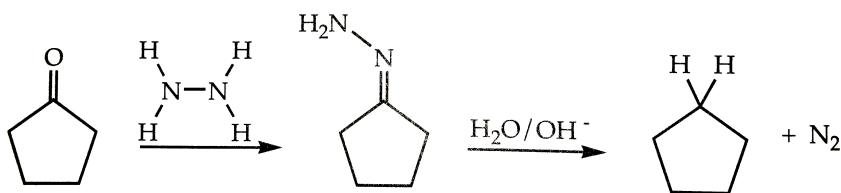


Figure 57: Example of Wolff-Kishner reaction carried out on cyclopentanone.

3. Reconnect the fragments with the proper chemical reagents.

Protecting groups are used when a reactant has multiple reactive sites, but you only wish to react at one of them. Aldehydes are converted into acetals, ketones are converted into ketals, and alcohols are converted into silyl ethers. You must be able to remove the protecting group at the conclusion of the reaction without invoking extreme conditions that may cause the molecule to react. As an example, Fig. 3.11 shows the conversion of an alcohol into a silyl ether and then back into an alcohol.

Acetoacetic ester has a carbon that is α to two carbonyl groups. It is referred to as a β -keto ester, because it has a

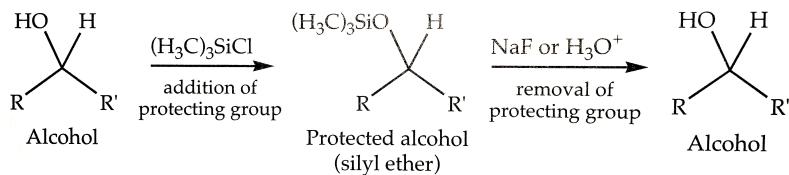


Figure 58: Protection and deprotection of an alcohol.

ketone group on the β carbon relative to the ester carbon. (β -keto esters are formed from a Claisen condensation reaction of two esters.) As a consequence, its protons are more acidic than typical α protons, so it can be deprotonated in mildly basic conditions, resulting in a carbanion. The carbanion is a strong nucleophile, capable of displacing a leaving group to form a new carbon-carbon bond. Fig. 3.11 shows the addition of an alkyl group to acetoacetic ester. Notice that the base of choice matches the alkoxy group of the ester, so that any transesterification reaction that takes place does not result in any change.

Malonic ester, like acetoacetic ester, has a carbon that is α to two carbonyl groups. As a consequence, the same

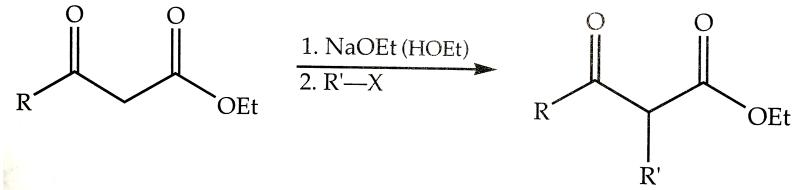


Figure 59: A typical reaction involving acetoacetic ester.

reactions carried out with acetoacetic ester can also be carried out with malonic ester. However, with malonic ester, a diester is formed, which has two functional groups that can undergo substitution, resulting in cyclization. This is shown in Fig. 3.11.

A β -keto acid (a carboxylic acid with a ketone group on the relative β carbon) can undergo **decarboxylation** when heated. This occurs readily, because the intramolecular hydrogen bonding of a β -ketoacid aligns the molecule for a shift in the π -electrons. Fig. 3.11 shows decarboxylation, the six-membered ring created by hydrogen bonding, and the decarboxylation products formed from that conformation.

Anabolism is a reductive process and the result is the build up of a molecule. Examples of anabolism include fatty acid synthesis and gluconeogenesis. **Catabolism** is an oxidative process and the result is the breakdown of a molecule. Examples of catabolism include glycolysis and β -oxidation. Typically, catabolism releases energy while anabolism requires, and thereby stores, energy. In humans, oxidations tend to take place in the *mitochondrial matrix* while reductions tend to take place in the *cytoplasm*, with glycolysis being a notable exception.

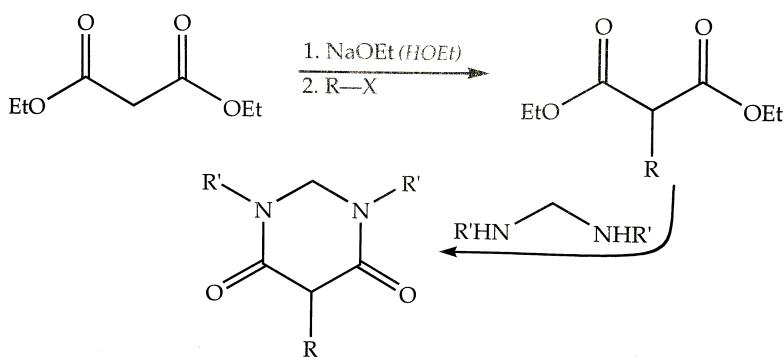


Figure 60: A typical reaction involving malonic ester.

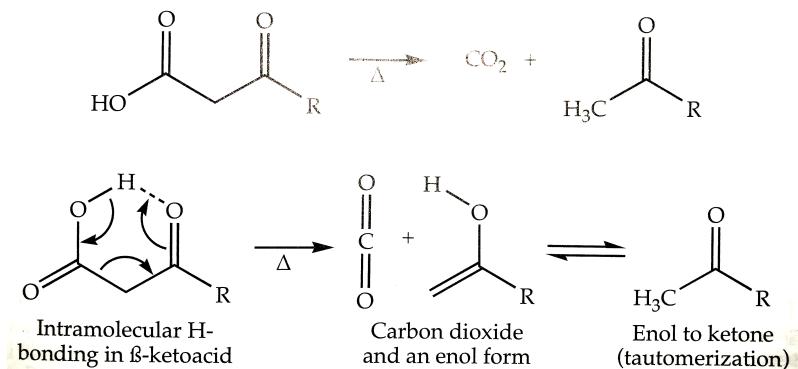


Figure 61: An example of decarboxylation.

3.12 Carbonyls and Alcohols—Key Points

Oxygen Containing Compounds

- Alcohols have higher boiling points than alkanes of roughly equal mass. The boiling points of alcohols are: primary > secondary > tertiary. The hydroxyl group can be converted into a better leaving group—they are *mesylated* or *tosylated* to form a good leaving group. SOCl_2 and PBr_3 convert an alcohol into an alkyl halide.
- Aldehydes have strong IR absorbances around 1725 cm^{-1} and ketones have strong IR absorbances around 1710 cm^{-1} . They readily form acetals and ketals with alcohol in acid and form hemiacetals and hemiketals with alcohols in base.
- Carboxylic acids form more reactive derivatives by converting the hydroxyl group into a better leaving group. SOCl_2 converts a carboxylic acid into an acyl chloride, heat converts two carboxylic acids into an acid anhydride, and an alcohol in acid converts a carboxylic acid into an ester. Carboxylic acids can be made by complete oxidation of a primary alcohol or aldehyde, or by hydrolysis of acid derivatives such as esters, acid anhydrides, acid halide, amides, and nitriles.
- Carboxylic acid derivatives include esters, lactones (cyclic esters), acid anhydrides, acid halides, and amides (where the hydroxyl group of a carboxylic acid is replaced by an amine).

Carbonyl Reactivity

- Carbonyl compounds are electrophiles that can be attacked at the carbonyl carbon by a nucleophile, forming a tetrahedral intermediate. Strong bases can also deprotonate carbonyl compounds at the α -carbon to form a nucleophilic C. They can also undergo oxidation-reduction chemistry. Recall that oxidizing agents are rich in O and reducing agents are rich in H

Name Reactions

1. Grignard reactions add one R-group to aldehydes and ketones, forming alcohols, and add twice to carbonyls with a leaving group, forming a tertiary alcohol.
2. Aldol condensation is the deprotonation of the α -carbon of an aldehyde or a ketone and the subsequent addition of the anion to the carbonyl carbon of either a second aldehyde or ketone. It forms an α,β -unsaturated ketone or an α,β -unsaturated aldehyde, and can exhibit kinetic versus thermodynamic preference with ketones. The Claisen condensation reaction is similar, but it converts esters into a β -ketoester.
3. Transesterification is the conversion of one ester into another by exchanging the alkoxy group, in either acidic or basic conditions.
4. Wittig reaction is the conversion of a ketone into an alkene via $\text{Ph}_3\text{P}=\text{CR}_2$.
5. Baeyer-Villiger reaction is the conversion of a ketone into an ester.
6. Iodoform reaction is the conversion of a methyl ketone into a carboxylic acid.
7. Wölff-Kishner reduction is the conversion of a ketone into an alkane in base.

Synthesis

1. The thought process for synthesis involves (1) changes, (2) retrosynthesis, and (3) choosing reagents. Protecting groups are used on reactants with multiple functional groups—OH group is protected as —OTMS and C=O group is protected as C(OR)₂.
2. Acetoacetic ester is a great nucleophile when the α -proton is deprotonated, so it readily attacks C=O. When treated with acid water and heat, it hydrolyzes and decarboxylates.
3. Malonic ester readily adds an R-group at the methylene carbon, and can cyclize with a reactant with two nucleophilic sites.
4. Decarboxylation is observed with β -keto acids or 1,3 diacids. It proceeds through a six-membered ring held together by an H-bond, and forms an enol which equilibrates with the carbonyl compound.

Carbonyl Biochemistry

1. Oxidation and reduction in biochemistry is referred to catabolism and anabolism, respectively. Catabolism occurs mostly in the mitochondrial matrix, while anabolism occurs mostly in the cytoplasm.
2. Biochemical reagents include NAD⁺/NADH (where NADH is the reduced form), FAD/FADH₂ (where FADH₂ is the reduced form), and acetyl coenzyme A (which picks up acetyl groups to form thioester).

3.13 Monosaccharides

Carbohydrates are organic compounds with the empirical formula CH₂O. Common sugars that you should recognize in all traditional structural representations (Fisher, Haworth, and chair) are glucose, ribose, fructose, galactose, and mannose. Monosaccharides of five carbons or more are typically found in cyclic structures rather than straight chain structures. To form the cyclic structure, a hydroxyl group attacks a carbonyl carbon to form either a hemiacetal (if the sugar is an aldose) or a hemiketal (if the sugar is a ketose).

α and β glycosidic linkages impact the reactivity and structural nature of polysaccharides. To cleave linkages, specific enzymes are required. Of note is *amylase*, the enzyme that cleaves α linkages in glycogen, amylose, and amylopectin, allowing us to break down starches. Humans lack the enzyme to cleave β -linkages between glucose residues, so they are unable to breakdown cellulose.

Monosaccharides have the basic formula C_nH_{2n}O_n, with one oxygen attached to every carbon and one unit of saturation, present as either a carbonyl group or a ring. Aldehyde sugars are **aldoses** and ketone sugars are **ketoses**. The length of the sugar is also considered in the name—four carbon sugars are **tetroses**, five carbon sugars are **pentoses**, and six carbon sugars are **hexoses**.

A monosaccharide is given D configuration if the hydroxyl group is to the right of the last stereocenter in a Fischer projection, whereas L configuration is given if the OH is to the left of the last stereocenter carbon. The numbering starts on

the side nearest the aldehyde or ketone group in linear monosaccharides. Another way of thinking about it is that the last chiral center (found on the penultimate carbon) in the D-isomers of the sugars has R-stereochemistry, while the last chiral center in the L-isomers of the sugars has S-stereochemistry. Note that there exists no correlation between D and L with (+) and (-) optical rotation. We are enzymatically programmed to digest D-sugars as opposed to L-sugars. As an example, take a look at the 4 stereoisomers of aldotetrose, shown in **Fig. 3.13**.

Sugars are routinely represented by **Fischer projections**, which are short hand for the top view (top perspective) of a

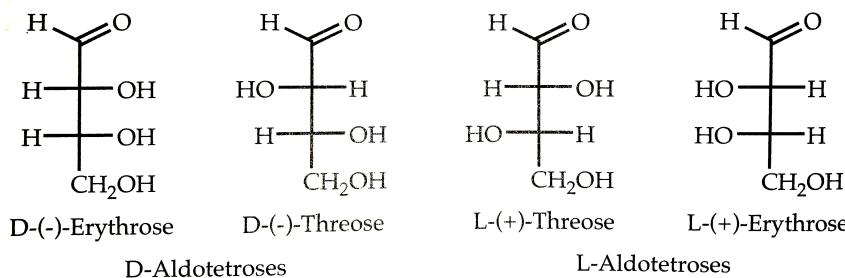


Figure 62: Fischer projections for each of the four stereoisomers of aldotetrose.

compound in its fully eclipsed conformation. The Fischer projection is derived from the straight chain form—any substituent on the right or left side of the main backbone is projecting out at you. The translation into a Fischer projection from a three dimensional structure is shown in **Fig. 3.13**.

Most sugars encountered in biology are pentoses and hexoses, as opposed to tetroses. Recall from stereochemistry that

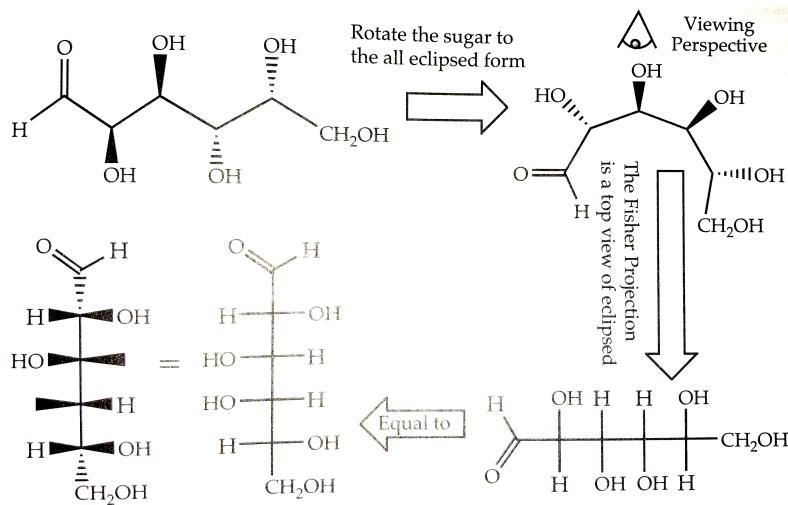


Figure 63: Example of translating a traditional chemical structure into a Fischer projection.

an **epimer** is one of a pair of stereoisomers that differ in the absolute configuration of a single stereocenter. When the molecule has only one stereocenter then the epimers are *enantiomers*. When the molecule has two or more stereocenters then the epimers are *diastereomers*. **Anomers** are cyclic monosaccharides or glycosides that are epimers, differing from each other in the configuration of C-1 if they are aldoses or in the configuration at C-2 if they are ketoses. The most important differences between them is this:

Anomers vary in chirality at the most oxidized carbon. Epimers vary in chirality at a less oxidized, backbone carbon.

(39)

The anomeric carbon is the former carbonyl carbon in a sugar. The most stable form of sugar is the cyclic form, which forms when a hydroxyl group attacks the carbonyl carbon to yield a hemiacetal. This generates a new chiral center at the anomeric carbon. Because the attack can occur from either top or bottom, there are two diasteromers formed. The two diasteromers

are referred to as anomers and the new chiral carbon is the anomeric carbon. The two anomers are referred to as the α (axial for D-pyranoses) and β (equatorial for D-pyranoses) anomers. The β anomer is typically more stable than the α anomer. It is only in the cyclic form (*furanose form* for five-membered rings and *pyranose form* for six-membered rings) that anomers are possible. This is because six-membered rings are referred to as **pyrans** while five-membered rings are referred to as **furans**. An example of anomers is shown in **Fig. 3.13**.

The best way to memorize the α vs. β distinction is that in **cyclic sugar structures**, the term α technically means

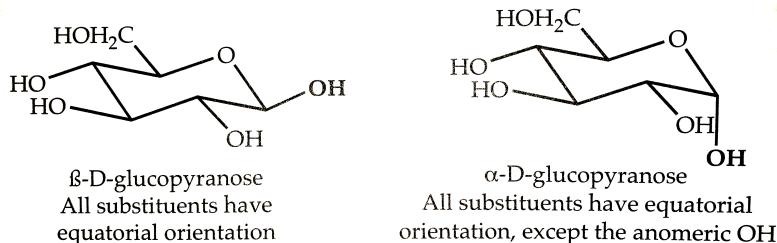


Figure 64: An example of anomers with α - and β - D-glucopyranose.

that the anomeric oxygen is trans with respect to the last carbon (C-6 for D-glucopyranose), and the term β technically means that the anomeric oxygen is cis with respect to the last carbon. The two anomers of a sugar exist in equilibrium *in vivo* and can interchange via a straight chain intermediate. Conversion from one anomer into the other is known as **mutarotation**.

Memorizing the names and structures for all sugars is no doubt a little pointless, but knowing the common monosaccharides is not a bad idea. Common monosaccharides in biochemistry include ribose, glucose, mannose, fructose, and galactose. To help recall common sugars, we shall present a mnemonic for three of them:

1. To help recall glucose, we have a not-so-nice way to remember the position of the hydroxyl groups. Using your right hand, if you ‘flip off’ the page with the fingernail of your middle finger facting the page, then your fingertips are positioned just as the hydroxyl groups of D-glucose are positioned in its Fischer projection. This helps with enantiomers too—D-glucose is found using your right hand while L-glucose is found using your left hand.
2. Mannose can be recognized by making a gun figurine using your middle and index fingers as the barrel (a ‘double-barrel’ gun)—your fingertips then mimic the positions of the hydroxyl groups in the Fischer projection of D-mannose. Statistically speaking, a man is more likely to have a gun than a woman, so mannose is based on a man gun hand.
3. D-Ribose has all of its hydroxyl groups on the right side of its Fischer projection, so ‘ribose is all right!’
4. Fructose is the ketose of glucose and galactose is the C-4 epimer of glucose.

These suggested mnemonics are shown in **Fig. 3.13**.

The **Haworth projection** is a simplified representation of cyclic sugars. They have no physical relevance, but are

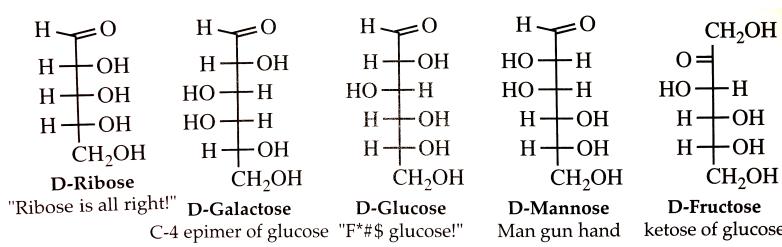


Figure 65: Mnemonic devices for memorizing the common monosaccharides.

easy to draw—they do not show the three-dimensional spacing that chair conformations show, but they emphasize positions above and below the ring, which is the major focus with sugars. Converting from the Fischer projection to the Haworth projection is easy once you get the hang of it. To go from straight chain to the ring form, recall the phrase ‘**downright uplefting**,’ which applies to carbons 2, 3, 4 in an aldohexose, but not the penultimate carbon (which has its oxygen in the

ring) or the anomeric carbon. Any hydroxyl groups on the right side in the Fischer projection end up below the ring in its cyclic structure, and any hydroxyl groups on the left in the Fischer projection end up above the ring in its cyclic structure. Numbering the carbons makes it easier to convert to the Haworth projection, because it helps to keep track of carbons 2, 3, and 4. The anomeric carbon is given the far right position in the Haworth projection. **Fig. 3.13** shows the generation of the Haworth projection of D-galactopyranose. An important skill to have is the ability to recognize sugars relative to glucose and to be able to quickly convert from the ring structure to the straight chain structure, as these are the common ways in which you will see sugars.

In addition to doing mutarotation, monosaccharides can convert from an aldose into a ketose via an *enediol intermediate*.

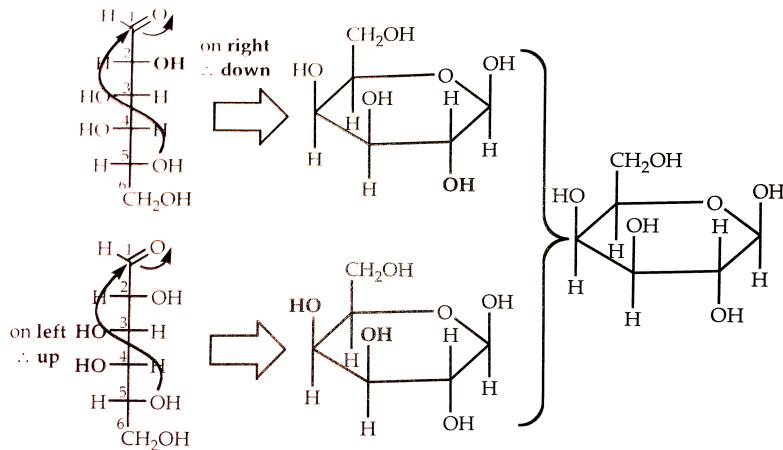


Figure 66: Generating the Haworth projection of D-galactopyranose.

in a conversion known as *ketose-aldehyde isomerization*. In glycolysis, the enzyme *isomerase* converts glucose-6-phosphate into fructose-6-phosphate via this process. Also in glycolysis, an isomerization process similar to tautomerization converts dihydroxyacetone phosphate into glyceraldehyde-3-phosphate. **Fig. 3.13** shows the step-by-step conversion of an aldose into a ketose in both acid and base.

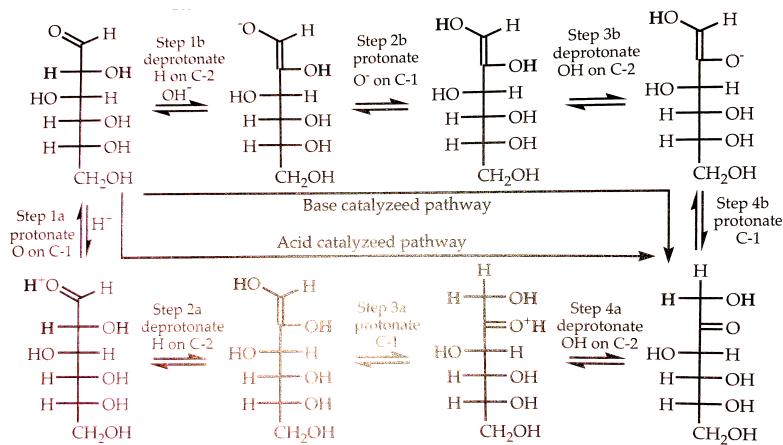


Figure 67: Mechanism of ketose-aldehyde isomerization. The bolded hydrogen atoms are the active hydrogens that are being either gained or lost in the reaction mechanism. The requirement for the conversion is the formation of either the enol (under acidic conditions) where the carbonyl oxygen is protonated and the α hydrogen is deprotonated, or the formation of the enolate (under basic conditions) where the α hydrogen is deprotonated.

3.14 Oligosaccharides and Polysaccharides

Sugars combine when a $-OH$ group of one sugar attaches the hemiacetal carbon of a second sugar, displacing the hydroxyl group from the hemiacetal carbon to form an acetal. This forms a **glycosidic linkage**, resulting from a dehydration reaction of the two hydroxyls involved. The terms used to describe the orientation of the anomeric hydroxyl group in a cyclic monosaccharide (i.e. α and β) are also used to describe the orientation of the oxygen in the glycosidic linkage—when the anomeric oxygen is cis with respect to the last carbon (i.e. C-6 in an alcohopyranose), the linkage is said to be β . The names for disaccharides are derived from the two component sugars where the left sugar (using hydroxyl 1 of an aldopyranose) is considered a substituent on the right sugar (using either hydroxyl 4 or 6). The left sugar gets an ‘-osyl’ suffix and the right sugar gets an ‘-oside’ suffix. For instance, lactose consists of galactopyranosyl and glucopyranoside. In most cases, the linkage is 1,4, as in the disaccharides maltose, cellobiose (formed from the hydrolysis of cellulose) and lactose. In sucrose, the linkage is 1,2, which is rare. **Fig. 3.14** shows the structures and nomenclature of these common disaccharides.

We store sugars in polymers such as *glycogen*, which has 1,4- α linkages between adjacent glucopyranoses and 1,6-

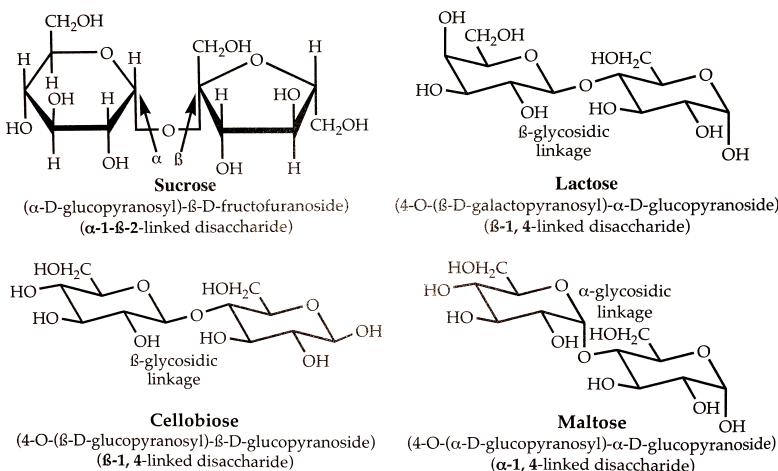


Figure 68: Common disaccharides and their structures.

branching about one out of every ten glucopyranose residues. Amylopectin and glycogen have branching, while amylose does not. The most abundant of all polysaccharides is cellulose (poly β -D-glucopyranose), which is used for structural purposes in plants. There are no 1,6-linkages found in cellulose, only 1,4-linkages. This means that cellulose is a linear polymer with no branching.

3.15 Chemical Reactions and Tests for Carbohydrates

Nitric Acid Sugar Oxidation: Nitric acid is rich in oxygen, making it an oxidizing agent. When an aldohexose is treated with nitric acid, both terminal carbons are oxidized into carboxylic acid groups. If the product, called an *aldaric acid*, is meso, then the compound is optically inactive. Using retro-analysis, the chirality of the hydroxyl groups in the original monosaccharide can be determined. **Fig. 3.15** shows the aldaric acid formation reactions of four aldohexoses.

Osazone Test: Really, the osazone test should be more aptly named the *epimer test*. When treated with three equivalents of phenyl hydrazine ($HPhNNH_2$), an aldose or 2-ketose undergoes a substitution reaction to form an imine, followed by the oxidation of an alcohol into a carbonyl, and lastly a second substitution reaction forming a second imine group. Imine moieties form at carbons 1 and 2, so any chirality at either carbon is lost. As a consequence, C-2 epimers yield the same osazone. **Fig. 3.15** shows that D-mannose, D-fructose, and D-glucose all yield the same osazone. The osazone test works in conjunction with other information to zero in on the identity of an unknown sugar.

Tollen's Test: The Tollen's test involves the reduction of Ag^+ by an oxidizable sugar. As silver ion is reduced to silver metal, if the solution is spun within the container, then silver precipitates on the inside walls, resulting in a ‘silver mirror.’ The Tollen’s test is said to be positive if the silver mirror forms rapidly, indicating the presence of an aldose.

Kilian-Fischer Synthesis: The Kilian-Fischer synthesis increases the carbon chain of an aldose by one. Note that it doesn't work with ketoses. The new carbon is added as a cyano group to the carbonyl carbon, so the chain increases by one carbon at the aldehyde end. After reduction to an imine and hydrolysis, the cyano group is converted into an aldehyde. The

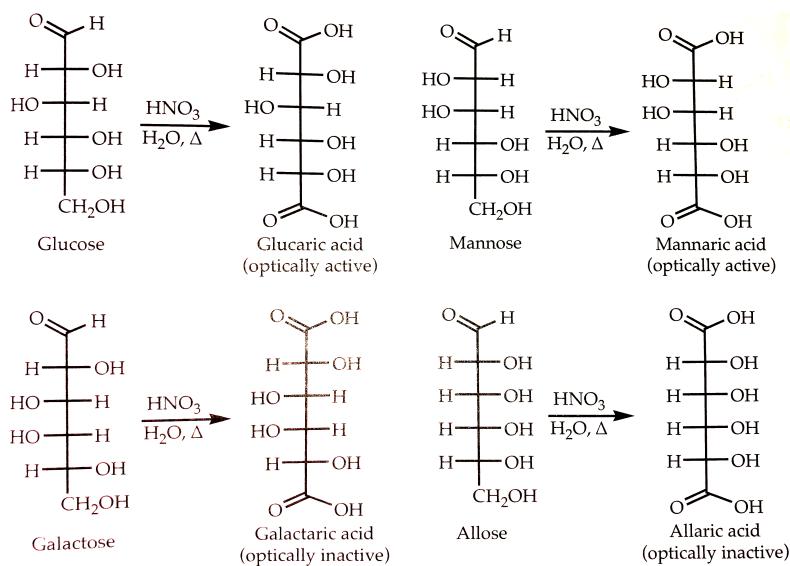


Figure 69: Nitric acid sugar oxidation test.

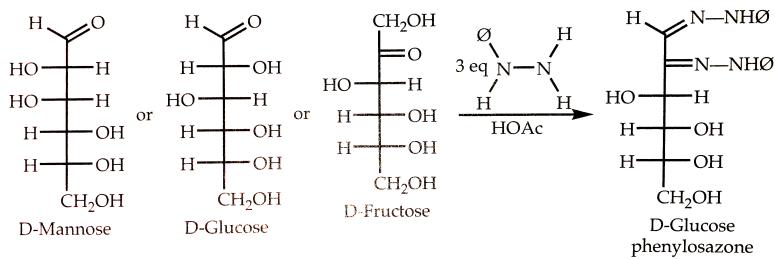


Figure 70: Example of the osazone test with C-2 epimers.

result is the formation of *two* aldose epimers with an additional carbon. As an example Fig. 3.15 shows the Kilian-Fischer synthesis of two D-aldohexoses from D-arabinose. An $-\text{OH}$ group in parenthesis represents the possibility of both chiral centers.

Ruff Degradation: In Ruff degradation, an aldose loses its first carbon, resulting in an aldose of one less carbon.

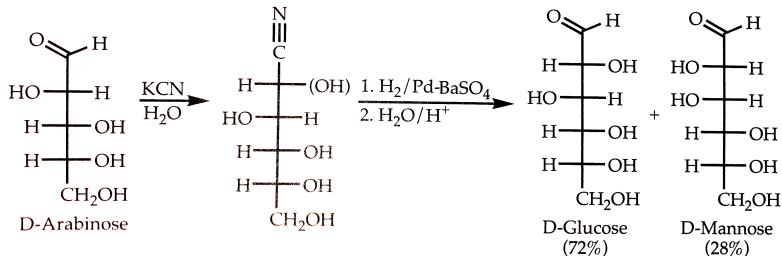


Figure 71: Kilian-Fischer synthesis of two D-aldohexoses from D-arabinose.

As we saw in the carbonyl section and in several biochemistry examples, losing one carbon is readily accomplished by decarboxylation. In order to decarboxylate the terminal carbon, it must be oxidized first to a carboxylic acid and then to carbon dioxide. Bromine in water effectively converts the aldehyde into a carboxylic acid (or in the ring form, a hemiacetal into a lactone). The carboxylic acid is oxidized into CO_2 by peroxide in the presence of a ferric cation catalyst. CO_2 leaves and decreases the chain length by one carbon. Fig. 3.15 shows the Ruff degradation of D-glucose. It is rarely used with aldopentoses, because when the product is an aldötetrose, the yield is low, likely due to the difficulty encountered with a

smaller ring in the cyclic form.

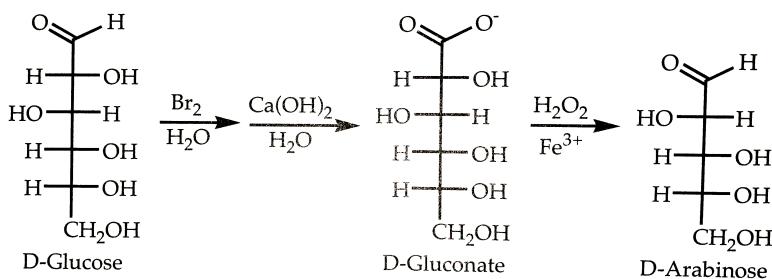


Figure 72: Ruff degradation of D-glucose into D-arabinose.

3.16 Biological Examples of Carbohydrates

As a general rule, natural sugars have D-chirality. This is because our enzymes recognize D-sugars, and therefore select for D-sugars. One notable exception to the ‘D only’ rule includes the presence of L-fucose, a reduced form of L-galactose, in the glycoproteins of blood type markers.

Carbohydrates are also important for **glycolysis**. Before we get into this, note that **you should *not* memorize glycolysis**, as the amount of time invested to do so would not have a good return in terms of increasing your MCAT performance. However, you *should* have a basic idea of what happens overall. You should know that glucose is converted into pyruvate and that there is a net generation of 2 ATP and a release of energy.

Glycolysis is a ten-step metabolic process. During the process, two net equivalents of ATP are formed, and also two equivalents of the oxidizing agent NAD^+ are also required. The basic steps are as follows:

1. Conversion of glucose into glucose-6-phosphate, basically tagging on a phosphate group to C-6. This is the first committed step in glycolysis, and releases a great deal of energy.
2. Step II involves isomerization of the aldose, and step III adds another phosphate group. Not super interesting.
3. In step IV, a six-carbon molecule is cleaved into 2 three-carbon molecules by a *reverse Aldol reaction* that breaks the bond between C-3 and C-4 of fructose-1,6-diphosphate (a β -hydroxyketone), forming dihydroxyacetone phosphate (DHAP) and glyceraldehyde-3-phosphate (G-3-P).
4. Step V converts DHAP into G-3-P by way of an isomerization reaction. As of step V, no oxidation has occurred.
5. Step VI oxidizes the aldehyde G-3-P into 1,3-bisphosphoglycerate (a carboxylic acid derivative). It is the *only* oxidation step in glycolysis.
6. Step VII dephosphorylates 1,3-bisphosphoglycerate into 3-phosphoglycerate, releasing a phosphate for the regeneration of ATP that was invested in earlier steps. Step VIII converts 3-phosphoglycerate into 2-phosphoglycerate. This is to position the phosphate for enol formation in the next step.
7. Step IX is enolate formation. Phosphoenolpyruvate is formed from 2-phosphoglycerate, and involves **dehydration**.
8. Step X finishes the process, converting phosphoenolpyruvate into pyruvate using the enzyme *pyruvate kinase*.

One pyruvate is formed, it can undergo either aerobic or anaerobic breakdown. Pyruvate can undergo fermentation, where C-2 is reduced from a ketone into an alcohol, resulting in the formation of lactate, which regenerates NAD^+ . We can’t reduce lactate any further (only yeast and bacteria can convert pyruvate into EtOH), so it is excreted as waste. Pyruvate can also undergo oxidative decarboxylation yielding CO_2 and Acetyl-CoA—this sets up the cell for performing downstream aerobic respiration.

3.17 Common Nitrogen Compounds

Unlike the drastic difference in reactivity between an ether and an alcohol, amines do *not* change their reactivity significantly as they become more substituted. Steric hindrance and electron density on the nitrogen are impacted by substitution, but not to the degree that tertiary amines are significantly more or less reactive than primary amines. Of significance when considering amines are their basicity and their nucleophilicity. This makes up the bulk of their reactions.

Amines are compounds with a central nitrogen that has three σ bonds to either C or H. The degree of substitution of the amine is described as *primary* (one substituent), *secondary* (two substituents), or *tertiary* (fully substituted with three substituents). If the nitrogen has a double bond to carbon, the compound is known as an **imine**. If a hydrogen on the amine nitrogen is replaced by an $-OH$ group, the compound is known as a **hydroxyamine**. Nitrogen when bonded to nitrogen by a σ bond is a **hydrazine**.

Amines are both weak bases and good nucleophiles. They are found in many neurobiology compounds because many amines have stimulating effects on the brain. Some examples include morphine, nicotine, cocaine, amphetamine, epinephrine, and mescaline.

Because the nitrogen of the amine has a long pair of electrons, nitrogen atoms can serve as a nucleophilic or basic center for a molecule. Amines are kind of complicated, as the impact of steric hindrance, hydrogen bonding in solution, and electronics is not as predictable as it is with most other compounds. In general, amines are often used as weak bases—the varying basicity of amines can be attributed to factors such as the inductive effect, steric hindrance, hydrogen bonding, and solvation. The seemingly chaotic order of relative amine basicity demonstrates the presence of both the inductive effect and steric factors. More specifically, steric factors would favor a smaller base, because of its ability to bind a proton with little to no hindrance. The inductive effect predicts that the more substituted amine is the most basic. The relative basicity of the methyl amines is $(H_3C)_2NH > H_3CNH_2 > (H_3C)_3N > NH_3$. As an example, consider the reaction $H_3CNH_2 + NH_4^+ \rightleftharpoons H_3NH_3^+ + NH_3$. Because the methyl group is *electron donating*, methylamine is more basic than ammonia. This means that the ammonium cation is more acidic than the methyl ammonium cation. Here is an example problem:

Given that the pK_b of NH_3 is 4.74, what is the approximate pK_b of Et_3N ?

- A -0.32
- B 3.25
- C 5.31
- D 9.26

Alkyl groups are electron donating, so triethyl amine Et_3N is more basic than ammonia. As a consequence, the pK_b for triethyl amine must be less than 4.74, eliminating choices C and D. A negative pK_b corresponds to a strong base and amines are weak bases, so choice A is too low. Therefore, choice **B** is the best answer.

Weak bases, such as amines, make good nucleophiles and the nucleophilicity of amines parallels their basicity for the most part. Deviations from this trend are due to steric hindrance. The majority of amine reactions in organic chemistry involve the amine behaving as a nucleophile by attacking an electrophile and either replacing a leaving group or breaking a π bond.

What is the major organic product when ammonia, NH_3 , is treated with excess ethyl iodide H_3CCH_2I following workup?

- A $H_2C=CH_2$
- B $H_3CCH_2NH_2$
- C $(H_3CCH_2)_2NH$
- D $(H_3CCH_2)_3N$

Because secondary amines are better nucleophiles than primary amines, which in turn are better nucleophiles than ammonia, the reaction proceeds until the alkyl halide is exhausted. With excess alkyl iodide, the reaction is free to add as many ethyl groups to the amine as possible. The best choice shows three alkyl groups added to the amine. This makes the correct answer choice **D**. It is possible that $(H_3CCH_2)_4N^+$, a quaternary amine, would form, but it is not listed as an answer choice.

When amines are added to a carbonyl, a nucleophilic attack at the carbonyl carbon takes place. If there is no leaving group on the carbonyl, then the reaction forms an **imine** and water, and is driven by its equilibrium. An **imine** is a carbon double bonded to a nitrogen ($C=N$). This reaction proceeds considerably better with an aldehyde than a ketone, attributed primarily to the greater electrophilicity of an aldehyde than a ketone. The basic idea here is that by saturating an aldehyde or ketone with a compound that contains at least two hydrogen atoms on the nitrogen, a $C=O$ bond is converted into $C=N$. With excess water, the reverse reaction is observed.

Just as ketones and aldehydes undergo tautomerization in the presence of mild acid or mild base to form an enol, imines also undergo **imine-enamine tautomerization** to form an **enamine**. Fig. 3.17 shows the equilibrium between an imine and an enamine. An imine can form from the reaction of an amine and a carbonyl group only if the nitrogen compound has two hydrogen atoms on the nitrogen (otherwise there's no way to form the water product). The formation of an enamine from a secondary amine and ketone is shown in Fig. 3.17.

Amides are carbonyl compounds with a nitrogen bonded to the carbonyl carbon. They are formed when an amine or

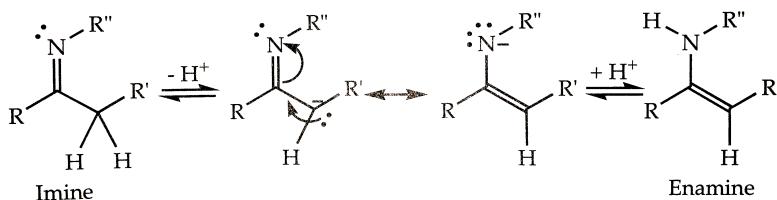


Figure 73: Mechanism of imide-enamine tautomerization.

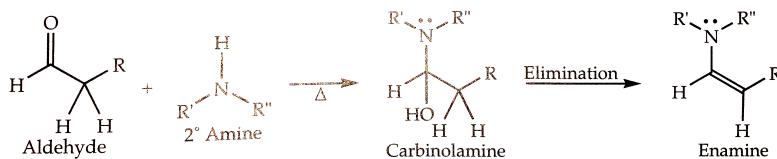


Figure 74: Formation of an enamine from a secondary amine and ketone.

ammonia is added to a carbonyl compound with a leaving group. In primary amides, the nitrogen is bonded to two hydrogens; in secondary amides, the nitrogen is bonded to one hydrogen; in tertiary amides, the nitrogen is bonded to no hydrogens. When an amide is part of a cyclic structure, then the compound is referred to as a **lactam**. Lactams are classified by the position of the carbon bonded to nitrogen relative to the carbonyl group of the amide. For instance, a β -lactam is a four membered lactam ring containing two sp^3 -hybridized carbons, a carbonyl carbon, and the amide nitrogen. Fig. 3.17 shows a β -lactam (a class of antibiotic), a γ -lactam, and a δ -lactam.

In an aqueous environment at neutral pH, the carboxyl terminal is deprotonated and the amino terminal is protonated in an amino acid, so it has at least two charged sites despite having no net charge—this is referred to as a **zwitterion**. However, in organic solvents and lipid environments they exist as an uncharged molecule. There are two synthetic pathways for making amino acids: the **Gabriel synthesis** and the **Strecker synthesis**. The Gabriel method involves reacting a phthalimide with malonic ester, followed by the addition of the R-group (side chain) and acid-catalyzed decarboxylation. The Strecker synthesis method involves reacting an aldehyde that contains the R-group of the desired amino acid, and the subsequent conversion into an imine, addition of a cyano group, and finally hydrolysis of the $C\equiv N$ triple bond using acidic water. Both of these syntheses are shown in Fig. 3.17.

3.18 Amino Acids and Proteins

Amino acids are essentially just polyprotic acids where all of the protons are weakly acidic. The amino terminal, when protonated, forms a cation with a pK_a of roughly 9.7 ± 0.9 , and the carboxylic acid terminal has a pK_a of roughly 2.2 ± 0.4 (these numbers are important, make sure to memorize them!). Amino acids also have a chiral center at the α -carbon in all amino acids except glycine. Naturally occurring amino acids have *S*-stereochemistry at C-2, although in biochemistry we refer to that as the L-amino acid. The only exception to this pseudo-rule where L corresponds to S is cysteine, where L-cysteine

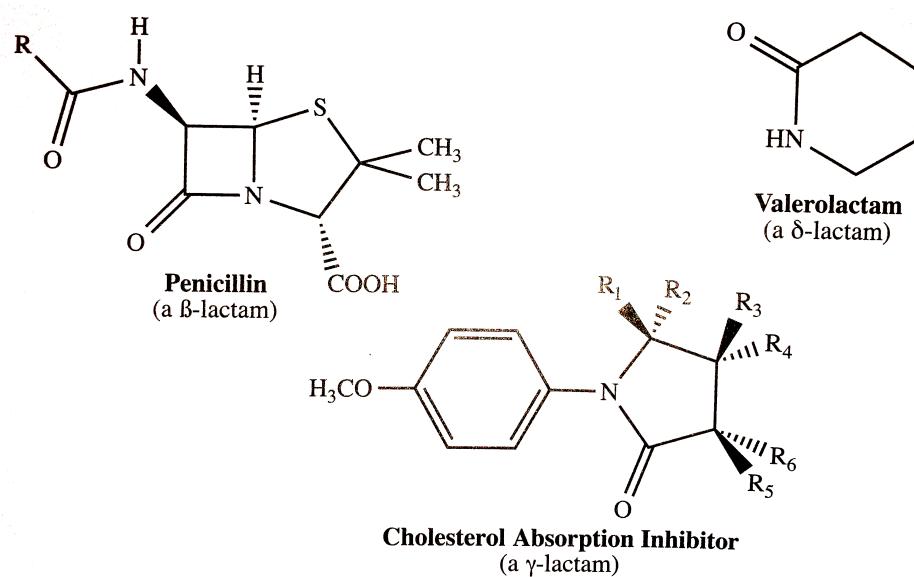


Figure 75: Examples of different types of lactams.

is the natural form, but it has R stereochemistry at the α -carbon. All of the amino acids and their structures/abbreviations are shown in Fig. 5.1—make sure to memorize it!

The hydrophobic amino acids are Phenylalanine (Phe, F), Leucine (Leu, L), Alanine (Ala, A), Isoleucine (Ile, I), and Valine (Val, V). In aqueous environments, hydrophobic amino acids are found in the core of a folded protein. *Semi-hydrophobic amino acids* are typically considered to be hydrophobic in most text books, but we segregate them from other hydrophobic amino acids because they don't pack as well as the alkyl side chains. These include the side chains with a nitrogen in a ring (i.e. Tryptophan (Trp, W), Proline (Pro, P)), a sulfur-containing Methionine (Met, M), and Glycine (Gly, G). Hydrophilic amino acids prefer to be situated in an aqueous environment—in water, hydrophilic amino acids are found on the exterior of a folded protein. These are sometimes referred to as *protic* side chains. Hydrophilic amino acids include Serine (Ser, S), Threonine (Thr, T), Cysteine (Cys, C), Tyrosine (Tyr, Y), Glutamine (Gln, Q), and Asparagine (Asn, N). A good mnemonic for recalling the six hydrophilic amino acids that don't fall into the acidic or basic category is “**See There, my CysTyr does not have a Glue Ass**” for remembering them in the order shown above.

Acidic amino acids have side chains that lose a proton from their neutral state. They have three acidic sites—the N-terminal, the C-terminal, and the side chain. In the case of Glutamic acid (Glu, E) and Aspartic acid (Asp, D), the side chain is a carboxylic acid, and so they are the two acidic amino acids. Cysteine and tyrosine can technically also lose a proton from their neutral side chain, but it happens at a high enough pH that we generally think of them as polar rather than as acidic. Basic amino acids have side chains that gain a proton from their neutral state. The three basic amino acids are Lysine (Lys, K), Arginine (Arg, R), and Histidine (His, H), and all have nitrogen-containing side chains. Like acidic amino acids, basic amino acids have their associated R-group pK_a values. A useful mnemonic is “**his lies are basic,**” which means that histidine, lysine, and arginine are basic. For histidine, the lower nitrogen in the imidazole ring is its basic site. Fig. 3.18 shows the charged amino acids in an aqueous environment at pH=7.

There are some fundamental structural features of proteins that result from properties of their component amino acids. Cross-linking of cysteine residues is one example. Another structural feature is *structural turns*, forced alterations in the structure of a protein, caused by proline. Proline is cyclic and thus unable to rotate freely about its σ bonds. It is locked into a cyclic conformation, which locks the protein backbone into a turn.

In aqueous solution at a pH of 7, an amino acid exists as an ionic species—the amino terminal is protonated and the carboxyl terminal is deprotonated. The net result is zero overall charge for hydrophobic amino acids, and the structure is referred to as a *zwitterion*. The pH at which the zwitterion exists in highest concentration is referred to as the **isoelectric pH**. The isoelectric pH is obtained by averaging the two pK_a values that involve the neutral molecule. The isoelectric point coincides with an equivalence point on a titration curve, the first equivalence point of all amino acids except the ones with basic side chains, where it coincides with the second equivalence point. To calculate the isoelectric pH, you must first decide if the side chain is acidic, basic, or neutral. For amino acids with neutral or acidic side chains, the isoelectric point is found

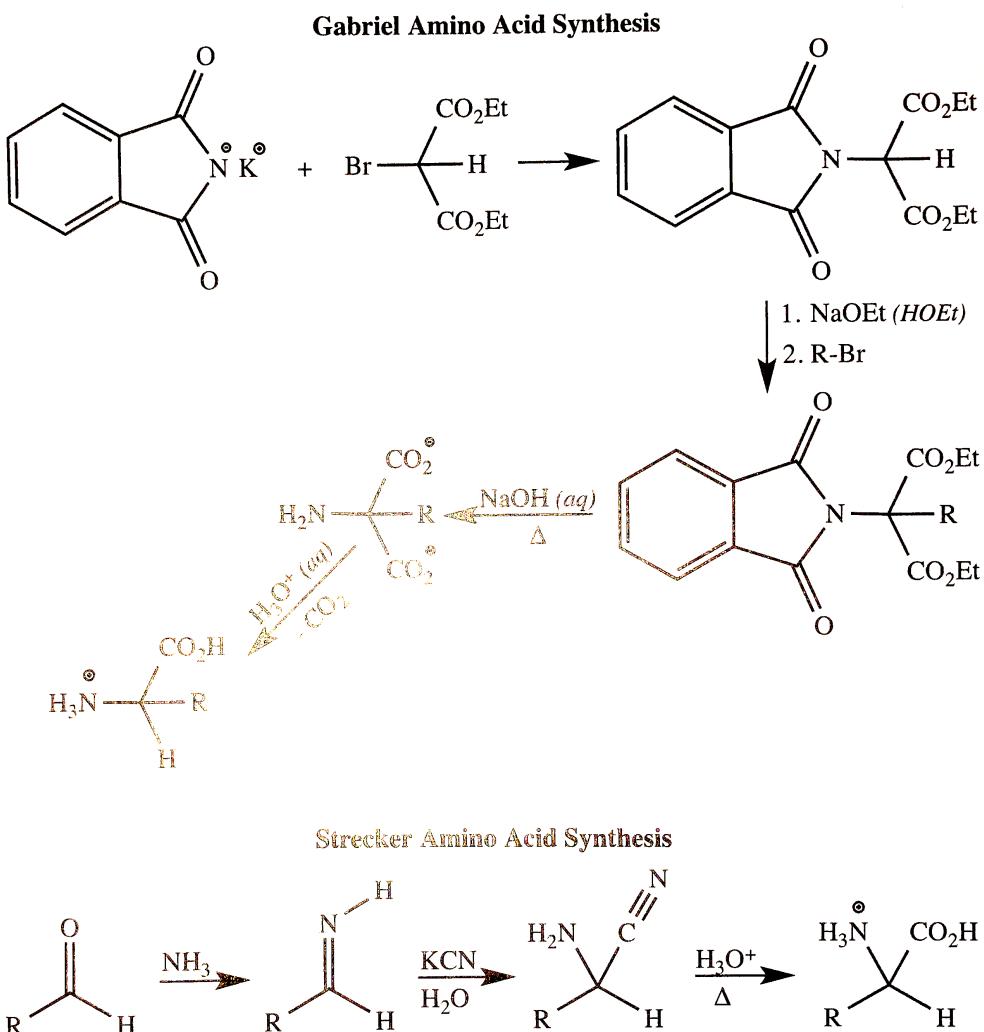
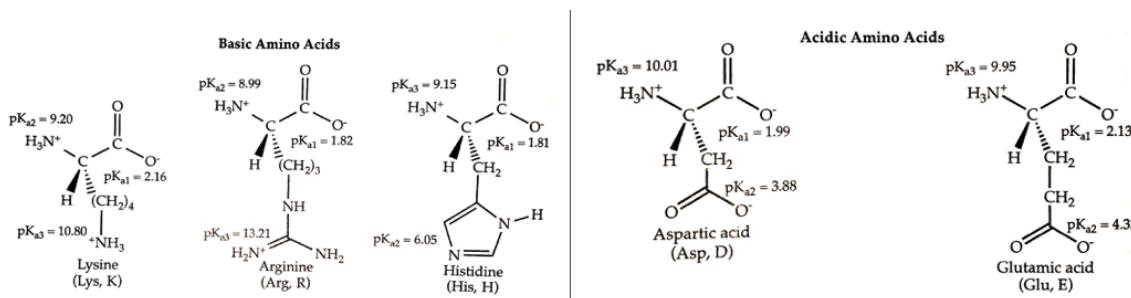


Figure 76: The Gabriel synthesis and Strecker synthesis of amino acids.

Figure 77: pK_a values and structures of charged amino acids in an aqueous environment at $pH=7$.

by averaging pK_{a1} and pK_{a2} . For amino acids with basic side chains, the isoelectric point is found by averaging pK_{a2} and pK_{a3} .

What if we want to get the isoelectric point pI of an entire protein? The pI for a protein is found by summing the number of basic amino acids plus one, and taking the pK_a with that numerical subscript and averaging it with the next pK_a value sequentially up. As an example, consider the short polypeptide chain shown in Fig. 3.18. This sample protein has a lysine

and histidine, so the pI is an average of pK_{a3} and pK_{a4} . The final step would be to determine which pK_a values are which. For this protein of interest, pK_{a1} is for the carboxyl terminal (this is true for *any* protein). The sequence of the remaining pK_a 's must be determined from numerical pK_a values.

The assignment of pK_a values is sequential with increasing numerical value, so the pI for this protein is found by aver-

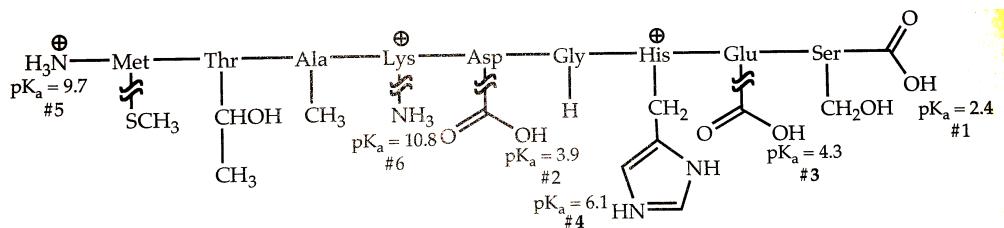


Figure 78: Example problem on how to determine the isoelectric point of a protein.

aging $pK_{a3}=4.3$ and $pK_{a4}=6.1$, which results in a pI of 5.2. This means that at pH=5.2, the protein exists as a zwitterion, at pH<5.2 it exists as a cation, and at pH>5.2 it exists as an anion. Here is another example problem:

What is the isoelectric point for tyrosine?

- A 2.2
- B 5.6
- C 6.2
- D 9.1

The isoelectric point is found by averaging pK_{a1} and pK_{a2} for all amino acids except Lys, Arg, and His. The isoelectric point for tyrosine is found using

$$pI = \frac{pK_{a1} + pK_{a2}}{2} = \frac{2.2 + 9.1}{2} = \frac{11.3}{2} \approx 5.6 \quad (40)$$

Therefore, choice **B** is the best answer. As a general rule, acidic amino acids have pI values *less than* 5.5, and basic amino acids have pI values *greater than* 7.5. All other amino acids have a pI value within the 5.5 to 7.5 range.

There are two types of gel electrophoresis lab techniques that you should know about. The first type of electrophoresis employs a pH gradient in the gel and the charged species migrate through the gel until they reach a pH equal to their pI. At that point, they have no net charge, so there is no electric force. This technique is referred to as **isoelectric focusing**. The second type of electrophoresis involves adding *sodium dodecyl sulfate*, SDS, to the protein mixture. The protein fragments incorporate SDS into their secondary structure, resulting in every fragment taking on a negative charge. Larger protein fragments can incorporate more SDS than smaller ones, resulting in a greater magnitude of charge. However, larger species are more massive. As a result, the incorporation of SDS creates an environment where all of the protein fragments have the same mass-to-charge ratio. Given that they are experiencing the same electric field, separation is due to size differences, where small species migrate faster than larger species.

Another technique that can separate proteins according to their charge is **affinity chromatography**. We will look specifically at **ion-exchange chromatography**, which separates by the affinity of charged proteins for oppositely charged sites on the polymer in the column (stationary phase). The stationary phase is made of an insoluble polymer, often polystyrene or cellulose, to which functional groups capable of carrying a charge have been attached. Positively charged columns bind anionic species and negatively charged columns bind cationic species. A protein carries a positive charge when the solution pH is less than the isoelectric point of the protein. When a mixture of proteins is added to the column and allowed to migrate down the polymer, selected proteins can be separated from the mixture when they bind the column, while proteins carrying the same charge as the column elute. To release a bound protein from an ion-exchange column, the polymer can be washed with a solution of varying pH or a solution of gradually increasing salt concentration. Note that both techniques for releasing the protein could denature it irreversibly, so the preferred technique is not always obvious.

There's also **size-exclusion chromatography** (also known as *gel-filtration chromatography*) to separate proteins. A

standard column is filled with agarose beads or dextran, which have natural pores of varying size. As the mixture migrates down the column, smaller species get caught within the pores, which slows down their progress, while larger proteins get caught in fewer pores, so they traverse the column in less time.

Determining the primary sequence of a protein is important, but automated sequencing has upper limits as to the number of amino acids that can be sequentially identified, so it is best to break a large protein into smaller, more manageable fragments. Each fragment can be sequenced and then the fragments can be reassembled to determine the overall sequence of the original protein.

The original protein must be marked such that the first and last fragments are easy to identify. The first amino acid in the sequence can be identified by adding 2,4-dinitrofluorobenzene (**Sanger's reagent**) to the protein, followed by complete hydrolysis using HCl. Sanger's reagent binds to the N-terminus of a protein, thereby labeling the first amino acid. The protein is then broken down into smaller fragments, often using chemicals/enzymes that are derived from nature.

Sequencing is often done in multiple steps. Prior to sequencing, a protein is typically treated with urea and β -mercaptoethanol. Urea denatures the protein by breaking the H-bonds holding it together, and β -mercaptoethanol breaks cross-linking. Once the protein's secondary and tertiary sequence is denatured, its primary structure can be sequenced.

To sequence the amino acids in a small polypeptide (20 to 50 amino acids typically), biochemists employ **Edman's reagent**, phenylisothiocyanate ($H_5C_6N=C=S$). This reagent sequentially removes each amino acid one at a time starting from the amino terminal. This leads to the automated sequencing equipment used in most biology labs.

A common rule in the chemistry of amino acids is that they exist exclusively as L-stereoisomers in biological systems. An exception to this "all amino acids are L" rule is found in the bacterial cell wall. The cell wall of a bacteria is held together by a net-like structure that involves a cross-link between D-alanine and glycine. As a point of biological interest, β -lactams such as penicillin act in a medicinal fashion by breaking this linkage and destroying the cell walls of bacteria.

4 Physics

4.1 Translational Motion

In general on the MCAT, although especially for the physics section, it's about speed more than precision. Important kinematic equations that you probably don't know yet:

$$\Delta x = \frac{1}{2} (v_0 + v_f) \Delta t, \quad v_f^2 = v_0^2 + 2a\Delta x \quad (41)$$

4.2 Forces and Torque

In order to handle formula identification problems *quickly* (as speed is extremely important on the exam), the two best techniques are **checking units** and considering **limiting cases**. For example, consider the following problem:

Example 2.3a The pulley system shown below, sometimes referred to as an *Atwood machine*, has two masses, m_1 and m_2 . Which of the following formulas represents the acceleration of this system?

A $a = \frac{(m_1 - m_2) g}{m_1 + m_2}$

B $a = \frac{(m_1 + m_2) g}{m_1 - m_2}$

C $a = \frac{g}{m_1 + m_2}$

D $a = \frac{m_2 g}{m_1 + m_2}$

This problem essentially shows two masses, m_1 , m_2 , attached on two sides of the pulley. There's a traditional solution to *actually* solve for the acceleration of the system, shown on pages 57-58 of TBR Physics, but this takes a lot of time. Instead, we can entirely avoid actually 'solving' by using the fact that the test is multiple choice. Choices C and D don't have the right units, so they are definitely wrong. Distinguish between A and B by considering limiting cases. If $m_1 = m_2$, then it

should be zero acceleration, not infinite acceleration, so thus, the right answer is **A**. See how quickly we were able to arrive at the right answer?

Recall that torque τ is defined to be $\tau = \mathbf{r} \times \mathbf{F}$. The length is sometimes called the *lever arm* or *moment arm*, and the center of the bolt is called the *pivot point*.

Although it may sound strange and possibly incorrect when you first consider the idea, it is possible for a static friction to accelerate an object from rest. There are two *everyday* examples that can help you to accept this concept: walking and driving. When you start walking from rest, you are clearly going from $v = 0$ to having a velocity in the direction you are moving. This means that you have accelerated. This is achieved by *pushing off* against the ground in a lateral direction. So, there must be a force accelerating you in a lateral direction. According to Newton's third law, there is an equal and opposite force as your push-off force. That opposing force is a static friction of the ground against your foot, as long as your foot does not slip. This explains why it is harder to start walking on an icy floor than a carpeted floor, because the static coefficient of friction μ_s for ice is much lower than μ_s for carpet.

In physics and engineering, **mechanical advantage** is the factor by which a machine multiplies the force put into it. The mechanical advantage of a lever/see-saw can be described by the following equation:

$$\text{mechanical advantage} = \frac{\text{weight of object}}{\text{applied force needed to support object}} = \frac{d_{\text{fulcrum}}}{d_{\text{weight}}} \quad (42)$$

where d_i is the distance from the fulcrum of object i . Example problems for force and torque:

9. Which of the following graphs (see Figure 4.2) best represents the force due to static friction as incline angle increases for a redwood block atop a smooth rubber surface?

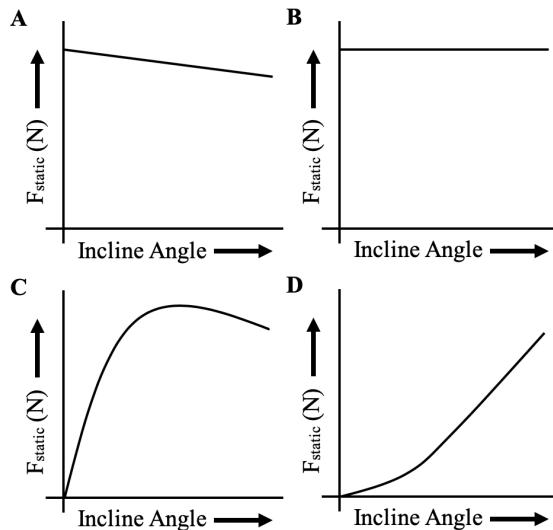


Figure 79: Answer choices for Question 9 (Taken from Physics Part I, Passage II, Question 9 in TBR)

Initially, you'd probably think that the answer is **A**; as you increase the incline angle, it makes sense the the static friction force would decrease, as this is what happens with the normal force, right? However, recall that static friction is a reactionary force, meaning that it is equal and opposite of any applied force. At zero inclination, the block has no inclination to move or anything like that, so the static friction is 0. As the block has more and more incentive to move as the incline angle increases (due to the force of gravity), the static friction must also increase in order to keep the block in place, until the object breaks free and begins to slide down the surface. Thus, the best answer choice in this instance is **D**.

22. If a person starts at the rim of a spinning platform and is pushed radially toward the central axis by a moving exterior wall, then what happens to the normal force felt by that person due to the wall?

- A It remains a constant.
- B It decreases, since r decreases.
- C It increases, since r decreases.
- D It decreases, since angular speed decreases.

By Newton's third law, we know that the normal force felt by the person is equal to the force the person exerts on the wall, which is basically equal to the centripetal force $F_c = v^2/r$. Superficially, this looks like an inverse relationship $F_c \propto 1/r$, and so we may naively guess the answer is C. However, as r decreases, v also changes for circular motion! Rather, only angular frequency $\omega = v/r$ remains constant, so thus, the form of F_c that we want to use is $F_c = (\omega r)^2/r = \omega^2 r$. From this equation, it is clear that in reality, $F_c \propto r$, and so the correct answer is B.

4.3 Work and Energy

When negative work is done *on* an object, it loses energy to its surroundings. Friction takes energy away from the sliding object. Generally, if *positive* work is done *on* the object, the object *gains* energy. If the object does positive work, the object *loses* energy. If the work is negative, then the opposite would be true. For example, the work done on a block by friction as it slides on a rough surface is negative and directly proportional to μ_k . The work done on a road surface, as a car skids across it, is positive and directly proportional to the skidding distance.

The three forms of energy you need to know are kinetic energy, gravitational potential energy, and spring potential energy. Power is the rate of doing work/transferring energy with respect to time.

Two simple machines that occasionally surface on the MCAT are the **inclined plane** and the **pulley system**. They make life easier, because they redirect and/or reduce the amount of *force* we need to apply in accomplishing some task. However, they do not lessen the overall amount of work that must be done! It still requires the same amount of energy to carry out the task. This is to say that because energy is conserved, the amount of work needed to perform an action will be constant, no matter how the action occurs. Other simple machines include **levers** (as we saw in the rotational equilibrium section) and **hydraulic lifts** (which we will see in the fluids section).

One way to characterize a machine is through its **mechanical advantage**, defined as

$$\text{mechanical advantage} = \frac{\text{weight of the object to be supported}}{\text{applied force needed to support the object}} \quad (43)$$

A machine with a bigger mechanical advantage will require less applied force than a machine with a smaller mechanical advantage when it is used to support some object. For pulleys in particular, note that the mechanical advantage is equal to the number of vertical cords supporting the moving pulley (page 111 in Physics Part I, TBR).

Heat and work values are defined from the system's perspective in chemistry. When heat q is positive, heat flows from the surroundings into the system. When q is negative, heat flows from the system out to the surroundings. When work w is positive, work is done on the system by the surroundings. When w is negative, work is done by the system on the surroundings.

The **Carnot cycle** converts either heat into work or work into heat. We inherently know the concept behind the Carnot cycle. When we blow on a hot liquid, it is done so with pursed lips. Consider blowing on the skin on the back of your hand. If you exhale through your mouth with a normal, relaxed degree of aperture, your breath comes out at body temperature. But, if you exhale through your mouth with a small opening, the air feels cooler as it passes across your skin. That is due to the compression of the gas (exothermic) as it passes through your lips, and the expansion of the gas (endothermic) once it leaves your mouth. The air feels cooler, because it is expanding as it passes across the surface of your skin. When a gas expands, the molecules increase their intermolecular distance, which breaks intermolecular forces. Just as bond breaking is endothermic, so is the expansion of a gas. The process of blowing air on your skin through pursed lips results in heat transfer from a cold body (your skin) to a hot body (your mouth). This is unnatural heat flow, so it is similar to the function of the Carnot heat pump. Another example of a Carnot cycle is any piston engine.

The *refrigerator* uses work to absorb heat as a fluid passes through the four stages in a closed system. The basic idea is to put work energy into the system to compress a gas and condense it into a liquid.

These are very complex topics. Just have the very fundamental perspective that a refrigerator takes in work (applied to a piston) and releases heat, while an engine takes in heat (to expand a gas in a piston) to release work. Do not overstudy this topic, even if you feel like you only partially comprehend it.

Let's look at a couple of problems about work and energy:

9. Consider a pulley system with a flat metal lift plate (with negligible mass) attached to one side and a counterweight of 50 kg attached to the other side. A box with mass of 90 kg is placed on the metal plate. What is true of the total work associated with raising the box compared the work needed to raise the counterweight and return the lift plate to the base position?

- A The work needed to raise the box exceeds the work needed to lift the counterweight.
- B The work needed to raise the box is less than the work needed to lift the counterweight.
- C The work needed to raise the box is equal to the work needed to lift the counterweight.
- D No work is required to raise the box.

Superficially, this seems like a very counterintuitive problem. Shouldn't it take no work at all to return the lift plate to the base? In actuality, the answer choices are simply talking about *work*—this work doesn't need to be done by a person, as it can also be done by gravity. Thus, the work (done by us) needed to raise the box is 40 kg times g (only 40 kg because of the counterweight), while the work (done by gravity) needed to restore the counterweight to its old position is 50 kg times g (50 kg because that is the mass of the counterweight). Therefore, the correct answer is **B**.

18. After a hand pump has been operating, the shaft is warm and the tip of the needle is cool. This is because:

- A air is compressed in the shaft of the pump, and it expands as it enters the needle tip.
- B air is compressed in the shaft of the pump, and it expands as it leaves the needle tip.
- C air expands in the shaft of the pump, and it is compressed as it enters the needle tip.
- D air expands in the shaft of the pump, and it is compressed as it leaves the needle tip.

We are given that the shaft is warm and the tip is cool. This means that the air is compressed in the shaft (because compressing the air is an exothermic processes where heat is released into the shaft) and expanded in the tip (because the gas molecules have more kinetic energy and absorb this energy from the environment, thus cooling the needle tip). This eliminates answer choices C and D. Just as we blow air out of our mouth, the air should be leaving the needle tip, not entering it, so thus, the correct answer choice is **B**.

4.4 Periodic Motion

The **crest** is the maximum of a wave and the **trough** is the minimum of a wave. Frequency and period for a cyclic process are inverses of one another and are related by the equation $f = 1/T$. If a system loses energy as it goes, we refer to this as *dampened harmonic oscillation*. The dampening is often caused by friction as the mass slides across the surface. The two

most important periodic motion set ups that you should know are the pendulum and spring:

$$\begin{aligned} F_{\text{pendulum}} &= -mg \sin \theta, & U_{\text{pendulum}} &= -mgL(1 - \cos \theta), & T_{\text{pendulum}} &= 2\pi \sqrt{\frac{L}{g}}, & f_{\text{pendulum}} &= \frac{1}{2\pi} \sqrt{\frac{g}{L}} \\ F_{\text{spring}} &= -kx, & U_{\text{spring}} &= \frac{1}{2}kx^2, & T_{\text{spring}} &= 2\pi \sqrt{\frac{m}{k}}, & f_{\text{spring}} &= \frac{1}{2\pi} \sqrt{\frac{k}{m}} \end{aligned} \quad (44)$$

Transverse waves cause vibrations perpendicular to the direction of wave propagation (i.e. the ‘wave’ at sports events), while longitudinal waves cause vibrations parallel to the direction of wave propagation (i.e. sound waves). The speed v of a wave is $v = \lambda f$. In the particular case of a transverse wave propagating along a string, the speed $v = \sqrt{T/\mu}$, where T is the tension in the string and μ is the mass per unit length.

A phenomenon known as a **beat** occurs when two waves of differing wavelengths interfere with each other. Generally, if waves of frequency f_1, f_2 interfere, they will give rise to a beat frequency of $f_{\text{beat}} = |f_2 - f_1|$. **Standing waves** are disturbances/waves that do not appear to travel along their line of propagation. Areas of no vibration are referred to as **nodes**. The area of the vibration where the amplitude is the largest is referred to as an **antinodes**. For the purposes of the MCAT, antinodes are found symmetrically displaced between adjacent nodes. The allowed standing-wave harmonic frequencies are given by $f_n = \frac{nv}{2L}$, where n is a natural number and v is the velocity of the wave. The allowed standing-wave frequencies are given by $\lambda_n = \frac{2L}{n}$.

The amplitude, and hence the energy of the system, may increase if we supply a vibrating force that has a frequency which is the same as the natural frequency of the system. This phenomenon is called **resonance**. The key to identifying resonance phenomena is to recognize that a given passage involves some object that can vibrate, an applied vibrational force, and a match between the natural and forced vibration frequencies.

Note that in basically all of the cases we are considering in the MCAT, fixed-end vibrating strings in a harmonic oscillation have the same wave speed at any point on the string.

4.5 Fluids and Fluid Statics

By definition, a *fluid* is any matter that flows. The **specific gravity** (also called the **relative density**) of a material is defined as the ratio of its density to the density of water at 4°C. That is,

$$\text{Specific gravity} = \rho_{\text{relative}} = \frac{\rho_{\text{material}}}{\rho_{\text{H}_2\text{O at } 4^\circ\text{C}}} \quad (45)$$

Pascal’s Principle states that a pressure applied to an enclosed fluid is transmitted equally throughout the fluid and to the walls of the fluid’s container. Mathematically, consider a column of liquid with pressure p_1 at height y_1 and pressure p_2 at height y_2 , with $y_1 < y_2$ and $p_1 > p_2$ since y_1 is lower than y_2 on the y -axis (which points upwards). Then, Pascal’s principle states that

$$p_1 - p_2 = \rho(y_2 - y_1)g \quad (46)$$

Let’s consider an example problem:

At a specific depth in a swimming pool, a barometer measures the total pressure to be twice that of atmospheric pressure. If the barometer is now submerged to a depth that is twice its initial depth, by how much does the total pressure increase?

- A The pressure increases by 50%.
- B The pressure increases by 100%.
- C The pressure increases by 200%.
- D The pressure increases by 300%.

To solve this problem, we can use the *ratio technique*. We have that, since pressure is additive,

$$p_{\text{total}} = p_{\text{atm}} + \rho gh \quad (47)$$

From the given information, we know that $\rho gh_{\text{initial}} = p_{\text{atm}}$, since the total pressure there is $2p_{\text{atm}}$. If the second depth is twice that of the first, then $\rho gh_{\text{final}} = 2p_{\text{atm}}$, meaning that the total pressure at the second depth is $3p_{\text{atm}}$. Taking the ratio of the final to the initial situation gives

$$\frac{p_{\text{total final}}}{p_{\text{total initial}}} = \frac{3p_{\text{atm}}}{2p_{\text{atm}}} = 1.5 \quad (48)$$

To complete this problem, just subtract 1 from the ratio to get the percentage change of +0.5 (i.e. 50%). Be careful of the wording in percentage questions. Here, we want a ‘percentage **change**’. If the question had asked instead, ‘What percentage of the initial total pressure **is** the final total pressure?’, the answer would have been ‘The final total pressure is 150% of the initial total pressure.’ Long story short, the best answer is choice **A**.

Archimedes’ Principle states that the buoyant force B experienced by an object is

$$B = \rho_{\text{fluid}} \cdot V_{\text{fluid displaced}} \cdot g \quad (49)$$

The following figure shows a summary of the mathematics associated with a floating object and a sunken object in a tank of fluid: Here is an example problem associated with the information shown in [Figure 4.5](#):

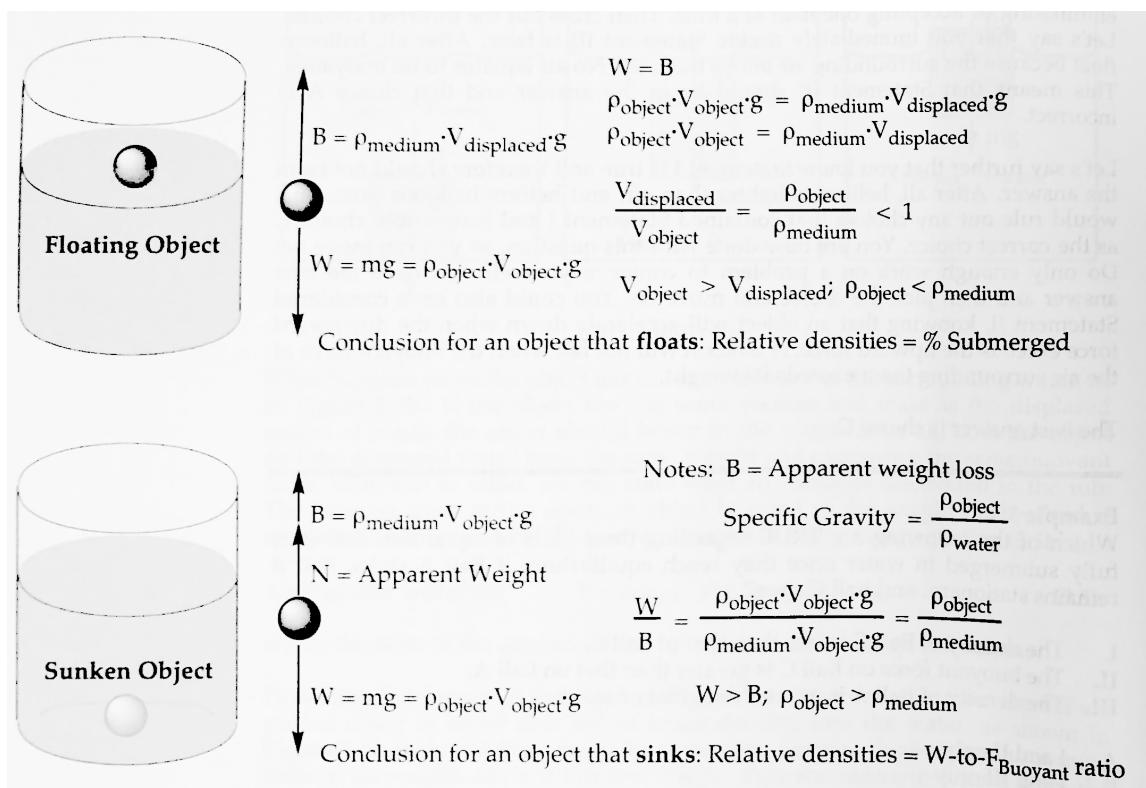


Figure 80: Mathematics of Objects placed in Tanks of Fluid

What is the specific gravity of an object that weighs 36N in air but only 9N in water at 4°C?

- A 0.25
- B 1.33
- C 3.00
- D 4.00

Because the object still has weight in water, this means that the object is a sunken object rather than a floating object. Thus, the object must be denser than the medium, so the specific gravity must be greater than 1. Based on the given information, the buoyant force must be 27N. Keep in mind that a sunken object rests on the bottom of the container, where because it experiences no acceleration, it has a net force of zero. This means that on the bottom of the tank, $mg = N + B$, where N is the perceived weight of the object in water, B is the buoyant force, and mg is the actual weight of the object in water. To get the specific gravity (the relative densities), we simply divide 36 by 27:

$$\text{Specific Gravity} = \frac{W}{B} = \frac{36\text{N}}{27\text{N}} = \frac{4}{3} = 1.33 \quad (50)$$

Thus, the correct answer is **B**.

The force associated with the tendency of the surface of a liquid to pull inward and shrink in area is known as **surface tension**. This is largely due to intermolecular forces such as hydrogen bonding or van der Waals interactions.

4.6 Viscosity and Poiseuille's Principle

Friction in fluids is called **viscosity** (η), which has units of $\text{N} \cdot \text{sec}/\text{m}^2$. Viscous forces retard the motion of one part of a fluid relative to another part of the fluid. If two layers of a fluid are held together tightly, then the viscosity of the fluid is said to be large, and the fluid will move *slowly*. Generally, when the temperature of a fluid *increases*, the viscosity of a liquid would **decrease** while the viscosity of a gas would **increase**. When a simple fluid flows through a pipe, the speed of that fluid is greatest at the central axis of the pipe and essentially zero along the walls of the pipe. As a fluid is flowing through the **center** of a pipe of length L with a certain speed v , there is a force that tends to oppose the fluid's motion. That force is called a **viscous retarding force** F and it is caused by the fluid's resistance to flow. The magnitude of F is given by

$$F = 4\pi\eta Lv \quad (51)$$

The net force on a fluid is proportional to the difference between the pressures of the pipe at the different ends. For the fluid to flow through the center of the pipe at constant speed, the driving force and the retarding force must be equal. If we substitute the cross sectional area $A = \pi r^2$ for a pipe, then we get

$$4\pi\eta Lv = A(p_1 - p_2) \longrightarrow v = \left(\frac{p_1 - p_2}{4\eta L} \right) r^2 \quad (52)$$

The **volume flow rate** Q is the volume V of fluid that passes through a pipe per unit time. Recall that we have mentioned that the fluid in a pipe flows the fastest at the center of the pipe and does not flow at all at the edges of the pipe. Using this assumption, we can express the average speed of the fluid as $\bar{v} = v/2$. If we also assume a circular cross-sectional area of $A = \pi r^2$, then we can arrive at $Q = V/t = (\bar{v}/2)\pi r^2 L$. Plugging in the result from earlier for the speed of fluid flow, we arrive at **Poiseuille's Principle**:

$$Q = \frac{\pi r^4}{8\eta L} (p_1 - p_2) \quad (53)$$

An important parameter in Poiseuille's Principle is the r^4 term. Note that if the radius of the pipe is doubled, the flow rate increases by a factor of 16. Although we have presented Poiseuille's Principle as a mathematical expression to this point, at the level of the MCAT it is highly conceptual. For instance, it explains that an extremely tall person will have a greater strain on their heart than a person of average height, because in order to generate the same volume flow rate Q , a tall person requires a greater Δp to offset the greater L associated with their longer circulatory system.

4.7 Continuity Equation

Assuming that fluid is incompressible, at any given moment in time, the fluid entering one region with cross sectional area A_1 must equal the fluid leaving the region at A_2 . In other words, the volume flow rates for the two regions must be equal. This gives us the **continuity equation**

$$Q = A_1 \bar{v}_1 = A_2 \bar{v}_2 \quad (54)$$

The continuity equation can also apply to branching as long as the fluid is incompressible and that the total cross-sectional areas of all the branches is considered. Here is an example of a problem involving the continuity equation:

Atherosclerosis is the thickening and hardening of the arterial walls. The aorta of a healthy adult has a radius of about 1.3 cm and a blood flow velocity of about 0.5 m/s. An atherosclerosis-stricken bacon epicure is found to have an effective aortic radius that is 75% of that in a healthy aorta. What is the blood flow velocity through the narrowed section of the artery? (Assume that the blood's volumetric flow rate is constant for all adults.)

- A 0.28 m/s
- B 0.38 m/s
- C 0.66 m/s
- D 0.89 m/s

This problem is easily solved using the continuity equation:

$$A_1 \bar{v}_1 = A_2 \bar{v}_2 \rightarrow \bar{v}_2 = \frac{A_1}{A_2} \bar{v}_1 = \frac{\pi (1.3 \text{ cm})^2}{\pi (0.75 \cdot 1.3 \text{ cm})^2} (0.5 \text{ m/s}) = \frac{16}{9} \cdot \frac{1}{2} \text{ m/s} = \frac{8}{9} \text{ m/s} \approx 0.89 \text{ m/s} \quad (55)$$

Therefore, the best answer choice is **D**.

4.8 Bernoulli's Equation

Bernoulli's Equation stems from the work-energy theorem as applied to a moving fluid system. It can be applied to a fluid that is incompressible, is frictionless, and has laminar flow. The equation is

$$P_1 + \frac{1}{2} \rho v_1^2 + \rho g y_1 = P_2 + \frac{1}{2} \rho v_2^2 + \rho g y_2 \quad (56)$$

All this equation is basically saying is that energy in fluids has to be conserved. This equation can explain some of the phenomena you may have observed in life. For example, consider the bottom of a shower curtain ‘attacking’ your ankles when you turn the water on. When you turn the water on, it falls from the shower head to the bottom of the shower, and as it falls, it picks up a good deal of speed. According to Bernoulli’s equation, since v goes up and y_1 barely changes, the pressure against the walls (in this case, the shower curtain) must drop. The water is fastest near the bottom of the shower curtain, so it is pushed towards your ankles by the higher air pressure outside the shower.

A common example of Bernoulli’s principle applied to engineering involves airplane wings. The top of the wing is curved in such a way that it is longer than the bottom, so if we compare the paths an air molecule would take during the period of time when the wing passes by, it must move faster to pass across the top of the wing than it would passing along the bottom of the wing. According to Bernoulli’s equation, a higher speed results in a lower pressure along the surface of the wing (since the gravitational potential energy term is *basically* the same for both above and below the wing). The result is a net force upward, which adds to the overall lift force on the wing.

4.9 Turbulence

When we looked at Poiseuille’s Principle, we say that adjacent layers of a fluid have the ability to slide past one another in a smooth and uniform fashion. This is called **laminar** flow. Poiseuille’s Principle only holds for laminar flow. If the flow of

a fluid is sufficiently high, then a chaotic and irregular pattern develops in the fluid—this is called **turbulent** flow. Viscous frictional forces increase during turbulent flow.

The **Reynolds number** N_R , defined by

$$N_R = \frac{2\rho\bar{v}R}{\eta} \quad (57)$$

tells us whether flow is laminar or turbulent, where ρ is the density of the fluid, η is the viscosity, \bar{v} is the average velocity of the fluid, and R is the radius of the vessel through which the fluid flows. $N_R < 2000$ means the flow is laminar, $N_R > 3000$ means the flow is turbulent, and anywhere in between means the flow is unstable. This means that flow can be either laminar or turbulent depending on the material. Here's an example of a problem involving turbulence:

Which of the following will decrease the chance of turbulent blood flow in a vein?

- A Widening the vein while maintaining the same flow speed.
- B Thinning the blood without changing its density.
- C Increasing the absolute pressure on each end of the vein by the same amount.
- D Lowering the blood density without thinning it.

Widening the vein increases R , which increases the Reynold's number and thus the chance of turbulent flow, so choice A is eliminated. Thinning the blood decreases η , which increases the Reynold's number and thus the chance of turbulent flow, so choice B is eliminated. Increasing the absolute pressure on each end of the vein doesn't do anything according to the mathematical definition of the Reynold's number, so choice C is eliminated. Lowering the blood density decreases ρ , which decreases the Reynold's number and thus the chance of turbulent flow, so choice D is the correct answer.

4.10 Electrostatics

The mass of a proton is about 1800 times greater than the mass of an electron, and the mass of a neutron is approximately equal to the sum of a proton and electron. Coulomb's Law is given by

$$F_e = k \frac{q_1 q_2}{r^2} \quad (58)$$

Field lines always originate on a positive charge and end on a negative charge. They represent the way a (+) charge would migrate. Equipotential lines and electric field lines are *perpendicular* to one another. The electrical potential energy of particles with like charges increases when the charges are brought towards each other. Conversely, the electrical potential energy decreases when opposite charges are brought towards each other. Opposite charges naturally attract; as they approach one another, their potential energy decreases. The reverse applies when two like charges are brought towards each other:

$$\Delta PE_q = k q_1 q_2 \left(\frac{1}{r_f} - \frac{1}{r_i} \right) = qV \quad (59)$$

The qV part applies when we want to calculate the change in electrical potential energy as a charged particle is moving through a voltage difference.

The **dielectric constant** κ is the ratio of the electrical force between two charges when they are in a vacuum compared to when they are in a medium. Mathematically,

$$F_{\text{medium}} = \frac{F_{\text{vacuum}}}{\kappa} \quad (60)$$

A strongly polar medium responds to an external electric field by reorienting its polar molecules, so that the net electric field within the medium becomes weaker than the external field. This occurs when the partial positive charges of the medium align near the negative plate and the partial negative charges align near the positive plate. Such a field strength reduction diminishes the interactive forces between all charges within the medium. Coulomb's law can be modified for dielectrics:

$$F_{\text{medium}} = \frac{F_{\text{vacuum}}}{\kappa} = \frac{k |q_1 q_2|}{\kappa r^2} \quad (61)$$

Note that when solid materials melt, their polarizability normally increases, meaning their dielectric constant increases. However, as temperature continues to increase, thermal motion subsequently reduces its polarizability because the high kinetic energy eventually surpasses the electrostatic interactions that led to the polarizability.

An **electric dipole** is established when two charges of opposite signs are separated from one another by some distance L . If p is defined to be the **dipole moment**, q is the charge, and L is the distance separating the charges $+q$ and $-q$, then we have

$$\mathbf{p} = q\mathbf{L} \quad (62)$$

by definition. The net force on an electric dipole in a uniform external electric field is zero. However, dipoles can experience a net torque. If we let the angle between the dipole axis and electric field be θ , then the torque created by the dipole couple can be quantified by

$$\tau = \mathbf{E} \times \mathbf{p} = qEL \sin \theta = pE \sin \theta \quad (63)$$

The torque is at its minimum when \mathbf{p} and \mathbf{E} are parallel (or antiparallel) to one another, and at its maximum when \mathbf{p} and \mathbf{E} are perpendicular to one another. The potential energy PE of the system can be given by

$$PE = -pE \cos \theta \quad (64)$$

Notice that the potential energy of the system is lowest (i.e. the system is most stable) when the dipole is parallel to the electric field, and highest (i.e. the system is least stable) when the dipole is antiparallel to the electric field.

4.11 Electromagnetism

Magnetic fields can exert a magnetic force F_{magnetic} on a charge moving in that field, known as the Lorentz force:

$$\mathbf{F}_{\text{magnetic}} = q\mathbf{v} \times \mathbf{B} = qvB \sin \theta \quad (65)$$

where θ is the angle between the velocity \mathbf{v} of the particle and the external magnetic field \mathbf{B} .

A **solenoid** is a helical winding of a conducting wire, such as a coil of copper wire wound on a cylindrical tube. The strength of the magnetic field B inside the solenoid is $B = \mu_0 nI$, where I is the current and n is the number of turns per unit length in the solenoid.

The combination of electric and magnetic forces can often be used to select for particles with certain velocities. Imagine a long chamber with an electric field $F_e = qE$ pointing down and a magnetic field $F_b = qvB$ pointing up (with the magnetic field pointing into the page). If we shoot particles with a distribution of velocities into the chamber, then particles that are too fast have $F_b > F_e$, meaning they deflect up, while particles that are too slow have $F_e > F_b$, meaning they deflect down. If $v = E/B$, then the particle travels straight and makes it through the chamber.

4.12 Electricity and Electric Currents

By definition, current flows in the same direction in which positive charges flow and in the direction opposite to that in which negative charges flow. Thus, current flows from regions of higher electrical potential to regions of lower electrical potential.

A **galvanic cell** converts chemical potential energy into an electromotive force (emf) that generates electrical flow. The **anode** is the source of electrons and is the negatively charged pole, while the **cathode** is the endpoint of electron flow and is the positively charged pole. The **resistivity** ρ of a material is a measure of how difficult it is for charges to conduct through the material; higher resistivities are associated with better electrically insulating materials. The resistivity relates to the current density J and the electric field E , as $\rho = E/J$. **Conductivity** $\sigma = 1/\rho$ is the reciprocal of resistivity. The larger the conductivity of a material, the better that material is at conducting current, so a material with low resistivity has high conductivity. The **resistance** R of a conductor is $R = \rho L/A$, where L is the length of the resistor and A is the cross sectional area of the resistor. In general, resistors oppose the flow of current, and thus drain energy/power from electrical circuits. **Power** P is the rate of conversion of electrical potential energy into some other type of energy:

$$P = IV = \frac{V^2}{R} = I^2 R \quad (66)$$

A **capacitor** is formed from two conductors separated by an insulator. The **capacitance** C of a capacitor is $C = q/V$; in other words, it is the amount of charge q that can be stored per volt of potential difference V across the two parallel plates. For the parallel plate capacitor, we usually have

$$C = \frac{\kappa}{4\pi k} \frac{A}{d} \quad (67)$$

where κ is the dielectric constant, A is the area of the plate, and d is the distance between the capacitor plates.

The MCAT only tests RC circuits that feature resistor(s) and capacitor(s) in series. **Kirchhoff's loop rule** states that the algebraic sum of the potential differences in any closed circuit loop is zero. **Kirchhoff's junction rule** states that the total current flowing through the pathways leaving the junction must equal the current that entered the junction. Resistors add in series and capacitors add in parallel. Consider the following two practice problems:

What is true of the voltage across, charge stored on, and energy stored by two unequal capacitors connected in parallel?

- A The larger capacitor has a smaller voltage gain and stores less energy; both capacitors store the same charge.
- B The larger capacitor stores more charge and more energy; both capacitors have the same voltage gain.
- C The larger capacitor stores more charge and has a smaller voltage gain; both capacitors store the same energy.
- D The larger capacitor stores less charge and has a larger voltage gain; both capacitors store the same energy.

When charge stores in a parallel capacitor circuit, the voltage drop across each capacitor is the same. Given that the voltage is the same, the larger capacitor will store more charge, since $Q = CV$. The larger capacitor will also store more energy, since $\Delta PE_q = \frac{1}{2}CV^2$. Therefore, the correct answer is **B**.

What is true of the voltage across, charge stored on, and energy stored by two unequal capacitors connected in series?

- A The larger capacitor has a smaller voltage gain and stores less energy; both capacitors store the same charge.
- B The larger capacitor stores more charge and more energy; both capacitors have the same voltage gain.
- C The larger capacitor stores more charge and has a smaller voltage gain; both capacitors store the same energy.
- D The larger capacitor stores less charge and has a larger voltage gain; both capacitors store the same energy.

When charge stores in a series capacitor circuit, the charge stored on each capacitor will be the same. Since the voltage drop across each capacitor is $V = Q/C$, this means that the larger capacitor will have a smaller voltage gain. The larger capacitor also stores less energy, since $\Delta PE_q = \frac{1}{2} \frac{Q^2}{C}$. Therefore, the correct answer is **A**.

To summarize, let's rehash the differences between parallel circuit elements and series circuit elements: Lastly, if we consider all the electrons of a conducting metal, and thus all of the electron levels shared between the metal atoms, we get what is known as *band theory*—electrons fill from the bottom levels up, so only a few electrons are actually in high enough energy levels to be able to transfer. The level up to which the electrons fill is called the *Fermi level*. For metals with low enough ionization energy, the electrons in and near the Fermi level are free to flow.

Series Circuit Elements

- Same current
- Different voltage drop
- Increased equivalent resistance
- Reduced equivalent capacitance
- Larger resistors drain more power
- $P = I^2 R$
- Devices are dependent of one another

Parallel Circuit Elements

- Different current
- Same voltage drop
- Reduced equivalent resistance
- Increased equivalent capacitance
- Smaller resistors drain more power
- $P = V^2 / R$
- Devices are independent of one another

4.13 Sound and Doppler Effect

Sound is an example of a compressional wave, so the speed with which particles can return to their original position following a disturbance in the medium dictates the speed of sound in that medium. The strength of the forces between the molecules in a given medium and the density of the particles in that medium determine how fast sound will travel in that medium. As a result, we see the general trend for speed of sound as $v_{\text{sound in solid}} > v_{\text{sound in liquid}} > v_{\text{sound in gas}}$. The speed of longitudinal sound waves in various ideal gaseous media can be calculated from the formula

$$v = \sqrt{\frac{\gamma P}{\rho}} = \sqrt{\frac{\gamma RT}{M}} \quad (68)$$

where ρ is the density of the gas, P is the pressure of the gas, γ is C_p/C_v , the ratio of the molar heat capacities at constant pressure and at constant volume, respectively, R is the ideal gas constant, T is the temperature, and M is the molecular mass. Conceptually, as a general rule regarding the speed of sound, we have

$$v \propto \sqrt{\frac{\text{Restoring Force or Molecular Kinetic Energy}}{\text{Molecular Inertia}}} \quad (69)$$

The intensity of sound dissipates with distance by a squared factor. The change in pitch of a sound from a high frequency as it approaches an observer to a low frequency as it moves away from the observer is called the **Doppler effect**. The relationship between the observed frequency f_L and the actual sender frequency f_s for a scenario where the listener is moving towards a stationary source is

$$f_L = \left(\frac{v + v_L}{v} \right) f_s \quad (70)$$

where v is the speed of the sound wave, and v_L is the speed of the listener. This equation essentially tells us that if you move towards the source of the sound (i.e. $v_L > 0$), then you hear a higher pitch, and if you move away from the source of the sound (i.e. $v_L < 0$), then you hear a lower pitch. Similarly, if the sound source is moving but the listener is stationary, we have

$$f_L = \left(\frac{v}{v - v_s} \right) f_s \quad (71)$$

This tells us that when the source of the sound move towards you, (i.e. $v_s > 0$), then you hear a higher pitch, and if the source of the sound moves away from you (i.e. $v_s < 0$), then you hear a lower pitch. Overall, combining these two gives us the overall **Doppler effect**:

$$f_L = \left(\frac{v \pm v_L}{v \pm v_s} \right) f_s \quad (72)$$

$v_L > 0$ if the listener is moving towards the source, and $v_L < 0$ if the listener is moving away from the source. $v_s > 0$ if the source is moving away from the listener, and $v_s < 0$ if the source is moving towards the listener.

A point along the string that is never displaced is called a **node**. A point along the string that reaches maximum displacement is called an **antinode**.

In an open pipe (where both ends are open), the standing wave frequencies are $f_n = n \left(\frac{v}{2L} \right)$, where $n \in \mathbb{N}$. In a closed

pipe (where one end is open and the other is closed). the standing wave frequencies are $f_n = n \left(\frac{v}{4L} \right)$, where $n \in \{1, 2, 5 \dots\}$. For many of the questions dealing with harmonic frequencies you will notice a pattern in the numbers. Look for differences between numbers or common denominators when dealing with harmonic frequencies of standing waves. If you can find a highest common denominator, then you have likely found the fundamental frequency. The **fundamental frequency** is the lowest frequency where standing waves is observed in a system.

When a wave moves faster than the standard speed of sound in that fluid, it said to be a **shock wave**. For example, this occurs when a plane flies faster than the speed of sound, referred to as ‘breaking the sound barrier.’ A shock wave typically has a large burst of energy followed by rapid dissipation of energy. The drop can be attributed to destructive interference between the shock wave and the expansion waves caused by the moving object. A shock wave’s initial burst of energy is typically observed as a rapid swell in the local pressure, which is known as the *sonic boom*.

4.14 Light and Radiation

Light in the 400 nm to 700 nm range can be detected by the retina and processed in the brain, leading to the phenomenon of color. Remember that in the presence of an external light source, an object will appear a certain color but then absorb the complementary color (which is why it appears as the particular color). The **photoelectric effect** basically says that an incident photon can cause the release of an electron, so long as the photon has an energy greater than the ionization limit of the material. The energy required to remove an electron from the surface of the solid material is referred to as the *work function*, ϕ . Energy in excess of the work function becomes kinetic energy for the electron that is emitted. Mathematically, the photoelectric effect is represented as

$$h\nu = \phi + \frac{1}{2}mv^2 \quad (73)$$

The **threshold frequency** corresponds to the photon equal in energy to the work function. Any photon with a frequency less than the threshold frequency does not have enough energy to ionize the material.

If unpolarized light is passed through a perfect polarizing filter, the vectors would radiate in only one plane, as the light wave has been linearly polarized. The polarizing filter blocks all light perpendicular to the polarizing axis, but it allows the light aligned with the polarizing axis to pass through.

When talking about interference, an important thing to bring up is Young’s double-slit experiment. If two slits are separated by a distance d , and are a distance L away from some screen with $L \gg d$, and light with a wavelength of λ is sent through both slits, then the distance y between two adjacent maxima of the interference pattern is

$$y = \frac{\lambda L}{d} \quad (74)$$

Here is an example problem involving interference patterns:

When laser light is sent through two slits, a periodic pattern of constructive and destructive spots appears on a distant viewing screen. Suppose the two slits are now replaced with a **diffraction grating** (essentially several slits, which can be assumed to be infinitesimally small). If the same total light intensity is used in both the double-slit and the diffraction grating experiments, then the bright spots of the interference pattern in the diffraction grating will be:

- A more intense and spaced together more closely than in the double-slit experiment.
- B more intense and spaced farther apart than in the double-slit experiment.
- C less intense and spaced together more closely than in the double-slit experiment.
- D less intense and spaced farther apart than in the double-slit experiment.

Without going into depth, we know that a fixed amount of light should reach the screen, so if the number of spots decreases, the intensity of each remaining spot should increase. This makes choices C and D highly unlikely answers. Similar to Young's double slit experiment with two slits, in the diffraction grating, the constructive spots will appear when the light rays from all slits are interfering as constructively as they can. Since the light rays emanating from several slits must now interfere in phase, the likelihood of a maximum is much smaller than it is in the double-slit experiment. The diffraction grating maxima will also be much narrower than the double-slit maxima. Thus, the maxima intensity must be greater for the diffraction grating experiment. Thus, the best answer is choice **B**. Here is another example problem:

A thin layer of plastic ($n = 1.46$) coats the surface of a glass plate ($n = 1.52$). When monochromatic light is shined normally onto the plastic from the air above, constructive interference occurs. Destructive interference will occur when the coated glass is immersed in which of the following liquids?

- I.** Water ($n = 1.33$)
 - II.** Carbon Tetrachloride ($n = 1.46$)
 - III.** Benzene ($n = 1.50$)
- (75)

A I only

B I and III only

C II and III only

D III only

The system involves light traveling from air ($n = 1$) to plastic ($n = 1.46$) to glass ($n = 1.52$). Phase flips of light occur when the light bounces off a medium whose refractive index *exceeds* that of the medium through which the light passes. In the original constructive set-up, light bounces off of an air-plastic interface and a plastic-glass interface. For both reflections, the refractive index of the reflecting medium is greater than that of the medium through which the light passes. For both reflections, the phase of the light will flip by 180° . To get destructive interference, we would need to change the experiment such that one of these two phase flips does not occur. If this is achieved, the single-phase flip would lead to destructive (i.e. 180° out-of-phase) interference between the two final rays. This happens only with benzene, so the correct answer choice is choice **D**.

This problem demonstrates the following: when a wave reflects from an optically denser medium, its phase flips 180° . Consider a film of oil on water, in air. The phase flips for both the wave that reflects off of the air-oil interface and the wave that reflects off of the oil-water interface. In both cases, the light bounces off a material whose refractive index n is higher than that of the material through which the wave moves. Thus, the phases of both waves flip, the effect cancels, and the waves interfere constructively. Now, if the oil film was replaced with a soap film, then only one of the two reflecting waves flips its phase. This single flip effectively adds half a wavelength to the path-length difference and results in destructive interference.

Diffraction is the distortion/bending of light waves when they pass through an opening or around an object. It is more significant when the wavelength is larger than or comparable to the opening the wave crosses. Diffraction is less significant or negligible when the wavelength is much smaller than the opening. This applies to all waves.

4.15 Geometrical Optics

The **law of reflection** says that the angle an incident light ray makes with a surface is equal to the angle that the reflected light ray makes with the surface. The index of refraction n is the ratio of the speed c of an EM wave in vacuum to the speed of that wave in a given medium. Mathematically, it is

$$n = \frac{c}{v} = \frac{v_{\text{vacuum}}}{v_{\text{medium}}} \quad (76)$$

Notice that $n \geq 1$ always. The frequency of light doesn't change as it passes through different mediums, so that means that the wavelength of the light must change to account for changes in the speed of the wave. Now, consider the setup shown in **Fig. 4.15**. Suppose that the indexes of refraction of two media are n_1 and n_2 , respectively. The incident light ray has

an angle θ_1 with respect to the normal as it strikes the water and an angle θ_2 as it passes into the water. The relationship between the index of refraction of each medium and the angle of refraction is governed by **Snell's Law**, which is

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 \quad (77)$$

Notice that if $n_2 > n_1$, the light ray will bend towards the normal, while if $n_2 < n_1$, the light ray will bend away from the normal.

A light ray bends away from the normal as it moves from a medium of higher refractive index to a medium of lower

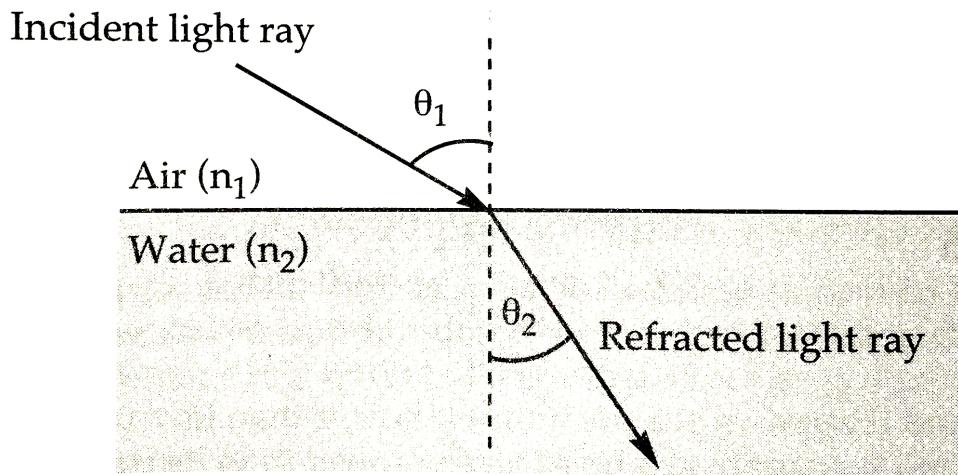


Figure 81: A demonstration setup for Snell's Law.

refractive index. If this refracted light ray were to keep bending away from the normal as the incident angle increases, it would eventually reach a refracted angle that is 90° away from the normal (meaning it is parallel to the surface of the medium). This incident angle is called the **critical angle** θ_c . At angles greater than θ_c , it would be totally *reflected* (not refracted) off of the surface of the water. This is called **total internal reflection**, which occurs since the sine of a refracted angle can never be greater than 1.00, so we can't satisfy Snell's law, meaning refraction can't occur. This is shown graphically in [Fig. 4.15](#). Here is an example problem involving total internal reflection:

A light ray traveling within an optically denser material is incident on an interface between two media. For which two indexes of refraction will its critical angle for total internal reflection be the SMALLEST?

1. When the indexes of refraction are 1.00 and 1.33.
2. When the indexes of refraction are 1.00 and 1.01.
3. When the indexes of refraction are 1.00 and 1.52.
4. When the indexes of refraction are 1.46 and 1.88.

Total internal reflection occurs only when light starts out in an optically denser medium (i.e. higher n index of refraction) and then hits an interface with an optically less dense medium. The critical angle comes from Snell's law. To get the critical angle, set the angle in the less dense (i.e. optically less dense) material to 90° . Solving this equation yields

$$\sin \theta_{\text{critical}} = \frac{n_{\text{less dense}}}{n_{\text{more dense}}} \quad (78)$$

The choice with the smallest ratio of indexes has the smallest critical angle. 1.00:1.52 is smaller than 1.00:1.01, 1.00:1.33, and 1.42:1.68, so thus, the best answer is choice C.

Electromagnetic waves of different wavelengths travel at different speeds through material substances (i.e. anything other

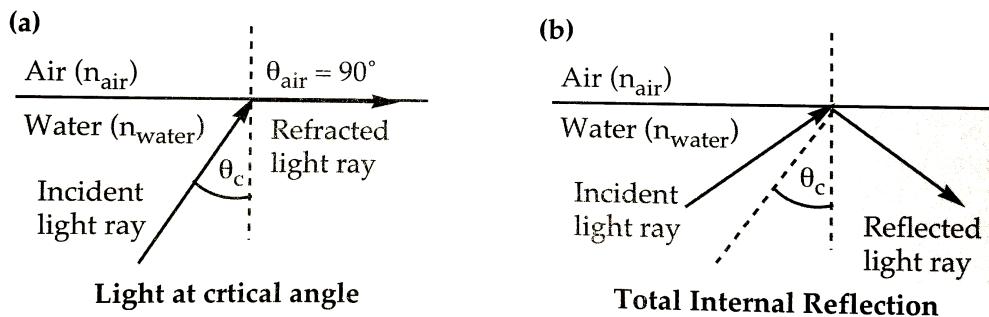


Figure 82: The critical angle and total internal reflection.

than a vacuum). As we have seen, the index of refraction for a given material depends on the wavelength of light passing through it. In other words, the amount that light bends at an interface between two materials varies with the wavelength. This gives rise to an effect known as **dispersion**. Note that red light (with large wavelength and hence fast speed through the medium) is deflected the least from its original direction of travel, while violet light (with small wavelength and hence slow speed through the medium) is deflected the most.

We now move on to **optics** which is basically just applied geometry. Because we are dealing with positions of objects and images, we will establish what in essence is a number line where the positive side represents the region of space where a real image can form and the negative side represents the region of space where a virtual image can form. A **real image** is one that can be projected onto a screen, while a virtual image cannot be cast onto a surface. We also have convention rules shown in the following table. We can relate the focal length os a lens or a spherical mirror with the distance separating

| Magnification | | Real/Virtual | |
|---------------|-----------|--------------|---------|
| Convention | Image | Convention | Image |
| $M < 0$ | Inverted | | |
| $M > 0$ | Upright | $i > 0$ | Real |
| $ M > 1$ | Enlarged | $i < 0$ | Virtual |
| $ M = 1$ | Same Size | | |
| $ M < 1$ | Reduced | | |

the object and image from the lens/mirror by the **Lens Maker's Equation**:

$$\frac{1}{f} = \frac{1}{s} + \frac{1}{s'} \quad (79)$$

where f is the **focal length** we're trying to solve for ($f > 0$ for converging lens/mirror and $f < 0$ for diverging lens/mirror), s is the **object distance** ($s > 0$ if object is to the left of the lens and $s < 0$ if object is to the right of the lens), and s' is the **image distance** ($i > 0$ if the image is real, $i < 0$ if the image is virtual). Using this, we can look at different types of mirrors. For plane mirrors (the flat ones in bathrooms), the image is always upright, virtual, and the same size. For convex (diverging) mirrors with $f < 0$, the image is always upright, virtual and smaller. For concave (converging) mirrors, the image can vary depending on the object and focal length. Regarding the focal length associated with a spherical mirror, its size relates to the radius of curvature of the mirror by

$$f = \frac{R}{2} \quad (80)$$

where R is the radius of the sphere that would be formed if the mirror were a complete sphere. We mentioned above that the image produced by a concave mirror is highly dependent on the object and focal length. We need to consider different cases of what can happen:

1. **Object is beyond the R of a converging mirror.** This means that the object distance is greater than $2f$. By the lens maker's equation, in order to satisfy it, the object distance is larger than R , so the image distance must be less than R in order for their reciprocals to add to $1/f$.

2. **Object at the R of a converging mirror.** This means that the object distance is equal to $2f$. This gives a real, inverted, and same size image.
3. **Object is inside the R , but beyond the f of a converging mirror.** This means that the object distance is between f and $2f$. This gives a real, inverted and larger image.
4. **Object is at f of a converging mirror.** No image will form.
5. **Object is inside of the f of a converging mirror.** A virtual, upright, and larger image will form.

Diagrams of each of these scenarios are shown in Fig. 4.15.

Here are a couple of practice problems regarding this:

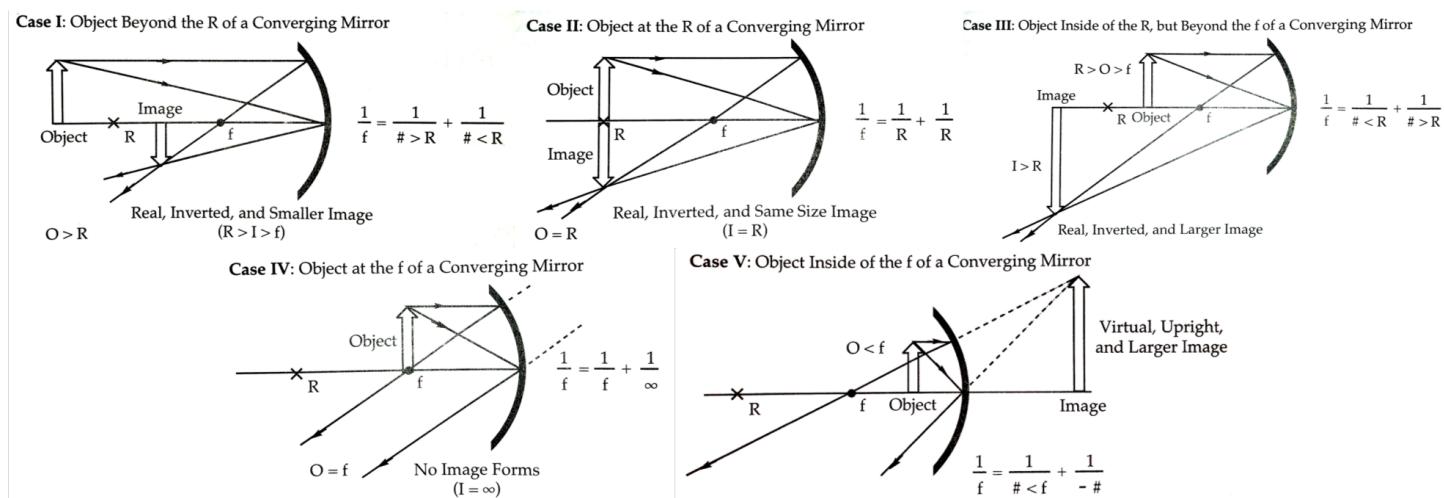


Figure 83: The scenarios and associated images of concave (converging) mirrors.

What is true of the image formed by a converging mirror with a focal length of 25 cm by an object that is 40 cm from the mirror?

1. It is an inverted, real image that forms 66.7 cm from the mirror.
2. It is an inverted, real image that forms 33.3 cm from the mirror.
3. It is an upright, virtual image that forms 33.3 cm from the mirror.
4. It is an upright, virtual image that forms 66.7 cm from the mirror.

The object is between f and $R = 2f$, so we should get a real image since further away from the mirror than the focal length, and also it should be larger. Thus, the answer is **A**.

What is true of the image formed by a convex mirror with a focal length of 20 cm by an object that is 30 cm from the mirror?

1. It is an inverted, real image that forms 60 cm from the mirror.
2. It is an inverted, real image that forms 12 cm from the mirror.
3. It is an upright, virtual image that forms 12 cm from the mirror.
4. It is an upright, virtual image that forms 60 cm from the mirror.

All convex mirrors produce virtual images, and for diverging mirrors, the image is always upright, forming inside of f . If $f = 20$ cm is given, then the image distance must be less than 20 cm. Thus, the answer is **C**.

Light rays coming from an object can be altered if they pass through a **lens**. A lens is an optical device with two curved, non-parallel refracting surfaces. There are two types of lenses to consider: converging and diverging. A **converging lens** (convex lens) with smooth surfaces allows parallel light rays to bend toward an axis that passes through the center of the lens and to converge at a single point called the **focal point f** . $f > 0$ for all converging lenses. Because light can enter the lens from both its right and left sides, the lens has two focal points. A **diverging lens** (concave lens) allows parallel light rays to move apart as they pass through the lens. The focal length of a lens is a **negative** value. This is because the diverging light rays appear to leave a single focal point from the side of the lens that light enters. Like converging lenses, diverging lenses will also have two focal points. In terms of real vs. virtual images, magnification, and image distance, the rules for single lenses are the same as they are for single mirrors, except that the real and virtual sides are flipped and convex and concave are reversed. As was the case with mirrors, diverging systems generate smaller, upright, virtual images, while converging systems, be it lenses or mirrors, also share similar rules.

Linear magnification m is the ratio of the height of the image h' relative to the height of the object h . These measurements are positive when upright and negative when inverted. Mathematically, we have

$$m = \frac{h'}{h} = \frac{s'}{s} \quad (81)$$

where we can also describe m using the ratio of image and object distances (s' and s , respectively) from the lens, since the height-to-length ratio ($h:s$) is the same for both the object and the image. The point of the MCAT is to be able to answer questions quickly. As such, here is a plan of attack for questions dealing with either a single lens or single mirror:

1. Identify whether the system is diverging or converging. Keep in mind that convex lenses and concave mirrors are converging systems, so you must do some work to determine whether the optic device is converging or diverging.
2. If the optic device is diverging, then the image will be smaller than the object, upright, virtual, and found between the focal length on the virtual side and the optic device. If the optic device is converging, then we need to do a positional analysis to see which case it fits.
3. Sketch a diagram of the object, optical device, and image. Double check using the lens maker's equation.

Two ideas to remember for a single lens or a single mirror are: **Upright images are always Virtual**, and **Inverted images are always Real**. Just remember **UV-IR**. The relationship works for any single lens or single mirror (whether it be convex, concave, or flat). Here are a couple of example problems:

Where is an image formed when an object is 20 cm to the left of a concave lens that has a focal length of 15 cm?

1. It is an upright, virtual image that is found 60 cm to the left of the lens.
2. It is an upright, virtual image that is found 8.7 cm to the left of the lens.
3. It is an inverted, real image that is found 8.7 cm to the right of the lens.
4. It is an inverted, real image that is found 60 cm to the right of the lens.

Let's use the three-step method. A concave lens is a diverging system, so the image will be upright, virtual, smaller than the object, and found between the focal length on the virtual side and the optic device. Thus, the best answer is choice **B**.

What is true of the image formed by an object 50 cm from a convex lens that has a radius of curvature equal to 60 cm?

1. It is an inverted, real image that is found 75 cm from the lens.
2. It is an inverted, real image that is found 30 cm from the lens.
3. It is an upright, virtual image that is found 30 cm from the lens.
4. It is an upright, virtual image that is found 75 cm from the lens.

Let's use the three steps. A convex lens is a converging system, so we need to determine which scenario it fits. The object is between R and f , so an inverted, real image will form beyond R . The image distance must be greater than 60 cm. So, the best answer is choice **A**.

Things get a bit more complicated when there are multiple lenses within an optical system. The basic rule to multiple lens systems is that the image of the first optical device is the object of the second optical device and so on for any additional lenses.

Real lenses can produce **aberrations**, owing to their shape and composition, so they do not always cast perfect images of an object. When lenses are not shaped properly, many types of aberrations can arise. There are three common types of aberrations:

1. **Spherical aberrations** develop when a misshapen lens still has cylindrical symmetry with respect to the optical axis, but has not been properly rounded; the image is clear when viewed through the center of the lens, but it becomes distorted as one moves the image off of the optical axis.
2. **Astigmatic aberrations** appear in a lens whose shape is neither properly round nor cylindrically symmetric. In this case, the image becomes blurred and distorted, regardless of the object's orientation. The two aforementioned aberrations can happen with any type of light passing through the lens, even monochromatic light.
3. **Chromatic aberrations** occur because of the differences in wavelengths for different colors of light. Recall that in a prism, light refracts differently depending on its wavelength. Since the focal point depends upon refraction through the lens, light of different colors refracts to different focal points. This is true for most lenses.

5 Biochemistry and Molecular Cell Biology

5.1 Amino Acids and Proteins

All of the standard 20 amino acids are referred to as *α -amino acids* (i.e. a 2-amino acid), except for *proline* which is referred to as an *α -imino acid*. Configurations (either relative or absolute) in amino acids refers to the stereochemical configuration around the chiral carbon. Due to differences in the priority of different amino acid side chains, not all amino acids have the same “absolute configuration,” which refers the R/S naming convention. However, all amino acids have the same “relative configuration,” which refers to the D/L naming convention. All biologically produced amino acids are in the L configuration.

The charged R groups are Asp (aspartate), Glu (glutamate), Lys (lysine), Arg (arginine), and His (histidine). All of these are highly ionized at neutral pH except for His, which is only weakly ionized.

Body fluids have pH range from 6.5 to 8.0—at these ranges, the amino and carboxyl groups are ionized. In other words, the α -amino group bears a positive charge, while the α -carboxyl group bears a negative charge. This is referred to as the **zwitterionic** nature of amino acids, due to the fact that the pKa of the α -amino terminal is about 9.4, while it is about 2.2 for the α -carboxyl terminal. Since amino acids can act as either an acid or a base, they are referred to as **ampholytes**. To determine the fraction of either the α -amino, α -carboxyl, or side chain groups that are ionized at a particular pH, use the **Henderson-Hasselbalch Equation**:

$$\text{pH} = \text{pK}_a + \log \frac{[\text{A}^-]}{[\text{HA}]} \quad (82)$$

The **isoelectric point** is the pH at which an amino acid carries **no net** electric charge. From the Henderson-Hasselbalch Equation, we know that the isoelectric point pI can be rewritten as

$$\text{pI} = \frac{\text{pK}_{a1} + \text{pK}_{a2}}{2} \quad (83)$$

Basically, it's just the main of all of the pKa values associated with the side groups for the molecule. For most amino acids, you only need to factor in the α -COOH and α -NH $_3^+$ pKa values, although the R side group may also contribute as well. In terms of making buffers, if a weak acid is within 1 pH unit of its pKa value, it resides within a good buffering range.

Polypeptides with a net positive charge at physiologic pH (~7.4) most likely contain amino acids with *basic* R groups. Polypeptides with a net negative charge at physiologic pH most likely contain amino acids with *acidic* R groups.

Electrophoretic separation of leucine from a protein sample would be least effective at pH 7.4, as opposed to pH values of 2.4, 1.4, and 0.4, because leucine has an aliphatic side chain, and so at physiological pH, leucine exists as a zwitterion.

Formation of **peptide bonds (peptide linkages)** requires an input of free energy; the hydrolysis of the peptide bond would be more favorable than its synthesis. Remember that peptide unit (i.e. the O–C–N–H bonding) is planar due to resonance!

The **central dogma** is essentially that DNA makes RNA makes protein. DNA makes additional DNA via DNA replication, DNA makes RNA via transcription, and RNA makes protein via translation; this has typically been the canonical view of how information flows in biological systems. However, there are other pathways as well—RNA can make cDNA via reverse transcription (common in retroviruses such as HIV), and there are also ncRNAs (noncoding RNAs) that can directly perform functions within the cell as an RNA molecule. Two examples of ncRNAs are tRNAs (transfer RNAs) and rRNAs (ribosomal RNAs), both of which are used for the translation of mRNAs into proteins.

There are a total of 20 canonical amino acids, shown in **Fig. 5.1**. Make sure to memorize all of them. There are four amino acids in particular that are worth talking about due to their individual special properties—histidine, proline, glycine, and cysteine.

1. Histidine is a special residue because its R group has a pKa of about 6.5, which is pretty close to the physiological pH of 7.4. Recall that at pH less than an amino acid's pKa, the amino acid exists in a protonated (positively charged form), and at pH greater than an amino acid's pKa, it will exist in the deprotonated form. Because histidine is in that 'special' regime, histidine is going to exist in both the protonated and deprotonated forms. So, this makes it a particularly useful amino acid to have at the active site of a protein where it can both stabilize or destabilize a substrate.
2. Proline has a secondary α amino group. All this basically means is that the side chain forms a second bond with the α nitrogen of this amino acid. Because of this unique cyclic structure of proline, this amino acid plays a central role in the formation of α helices and β sheets secondary structures. More specifically, it disrupts α helices because of the secondary α amino group, thus introducing kinks into α helices. Proline is thus an α helix breaker.
3. Glycine has just one hydrogen atom as its side chain. Because of this, the central carbon is now achiral (the only amino acid to be non-optically active). Additionally, because of the small R group, glycine is very flexible, so there's a lot of free rotation around the α carbon. Glycine also disrupts α helices because of its enhanced flexibility, thus also introducing kinks into α helices, and is thus also an α helix breaker.
4. Cysteine has a special thiol –SH group as its R group, meaning that when two cysteine residues are in close proximity, then their side chains can form an S–S bond called a *disulfide bridge*. Disulfide bridges exist in oxidizing environments, because the normal thiol groups exists in the reduced form in a reducing environment. For example, the extracellular space is an oxidizing environment, so this will favor the formation of disulfide bridges. However, in the cytosol (which is a reducing environment), disulfide bridges will likely not form.

5.2 Carbohydrates & Lipids

Carbohydrates have an empirical formula $(\text{CH}_2\text{O})_n$. They also must have an *aldehyde* or *ketone* functional group and at least *two* alcohol functional groups. If a carbohydrate contains an aldehyde, it is referred to as an **aldose**. If it contains a ketone, it is referred to as a **ketose**.

5.3 D- and L- Isomers

In Fischer projections, we care about the chiral carbon that is most distant from the carbonyl carbon. This chiral carbon is referred to as the **reference carbon**. If the hydroxyl group attached to that reference carbon is to the right, the molecule is the **D isomer**. If the hydroxyl group is to the left, the molecule is the **L isomer**. Most of the naturally occurring sugars are found in their D form (while most of the naturally occurring amino acids are found in their L form).

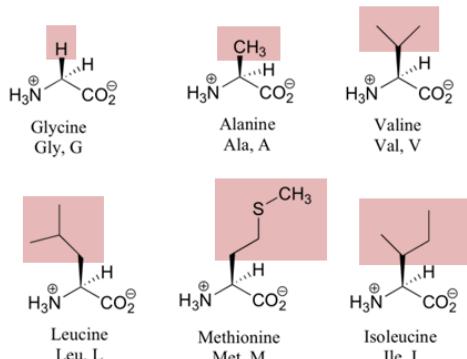
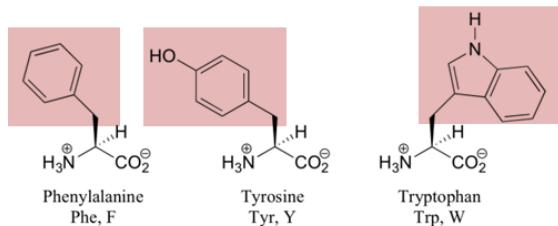
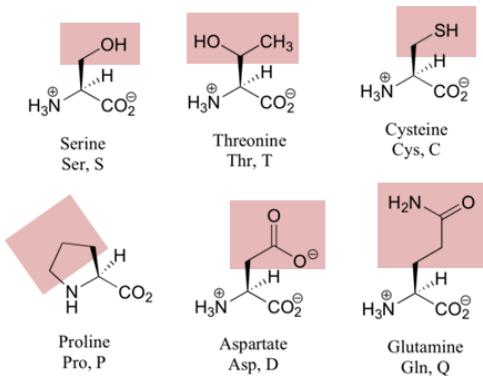
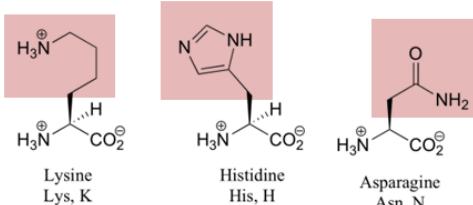
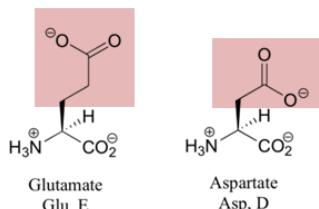
Nonpolar, aliphatic side groups**Aromatic side groups****Polar, uncharged side groups****Positively charged side groups****Negatively charged side groups**

Figure 84: 20 Canonical Amino Acids, all with L configuration

5.4 Monosaccharides, Oligosaccharides, and Polysaccharides

Sugars with five or more carbon atoms in their backbone prefer to be in the cyclic form. The C-1 carbon is the carbonyl carbon, also called the **anomeric carbon**. Two diastereomers of cyclic monosaccharides are α and β . In the α -anomer, the $-OH$ group at the anomeric carbon is on the *opposite* side of the ring from the $-CH_2OH$ group that is attached to the reference carbon. In the β -anomer, the $-OH$ group at the anomeric carbon is on the *same* side of the ring as the $-CH_2OH$ group that is attached to the reference carbon. An example of anomers is shown in Fig. 5.4. Notice that, depending on the starting sugar, formation of the cyclic ring creates either a *hemiacetal* group or a *hemiketal* group. Two oxidizing agents used to identify the functional groups of carbohydrates are **Tollens' reagent** (which contains Ag^+) and **Benedict's reagent** (which contains Cu^{2+}). If an aldose or ketose is capable of reducing these ions, those sugars are referred to as *reducing sugars*. Reducing sugars have the suffix **-ose**, while non-reducing sugars have the suffix **-ide**. Carbohydrates that contain a hemiacetal or a hemiketal group give positive tests with Tollens' and Benedict's reagents.

A sugar is a non-reducing sugar if its hemiacetal or hemiketal group has been converted to an acetal or ketal group, respectively (i.e. through reacting with an alcohol). These non-reducing sugars will not react with either the Tollens' reagent or Benedict's reagent.

The structure of lactose disaccharide has been given on the MCAT a number of times. Two important storage polysaccharides are *starch* and *glycogen*. Starch is a food reserve in plants. It is basically a bunch of D-glucose molecules linked together primarily with $\alpha(1 \rightarrow 4)$ linkages, although there can be other types depending on the type of starch. Glycogen is the storage polysaccharide common to all animals, and is located primarily in skeletal muscle and liver tissue.

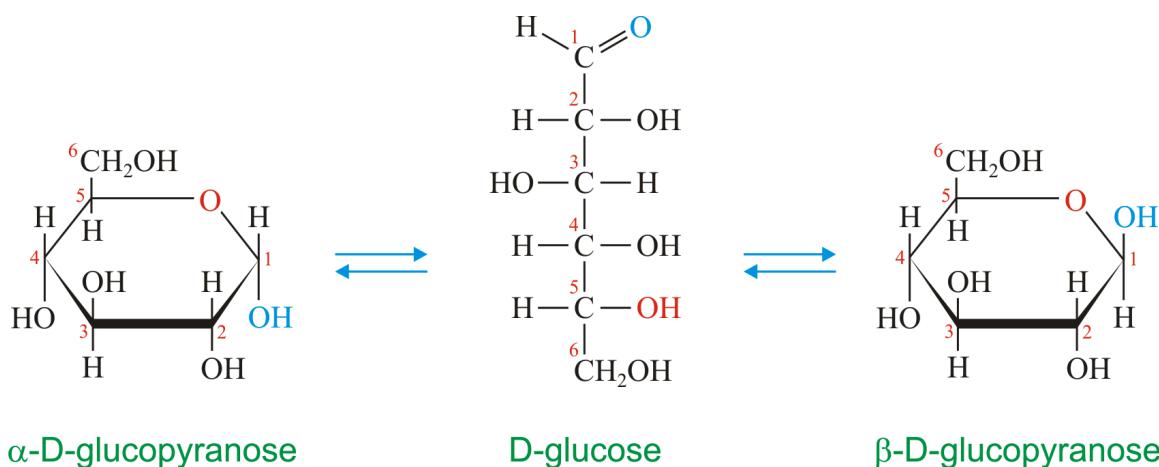


Figure 85: Formation of the two anomers of D-Glucopyranose

5.5 Lipids

Fatty acids are carboxylic acids with a hydrocarbon side chain. They can be either saturated or unsaturated. In nature, fatty acids are rarely free. Rather, they are esterified to a glycerol backbone to form a *triacylglycerol*. There are a number of different types of lipids that we will discuss here:

1. **Glycerophospholipids** (or phosphoglycerides) are the ones that we're most familiar with—two fatty acids esterified to the C-1 and C-2 carbons of glycerol, and a phosphate group attached to C-3. These molecules are amphiphilic, as they have nonpolar tails and a polar head.
2. **Sphingolipids** do not have a glycerol backbone, and instead are derivatives of amino alcohols. The C-2 carbon has an amino group, and a fatty acid can be attached to it as well through an amide linkage—this molecule is called a *ceramide*. Sphingomyelins are special types of sphingolipids that have a phosphoethanolamine or phosphocholine group attached to the C-1 carbon of the ceramide, and are abundant in the myelin sheaths that surround the axons of nerve cells.
3. **Cholesterols** are synthesized in the cytosol. Steroids are based on cholesterol and synthesized in the mitochondria. Cholesterol is transported into the mitochondrion, and adrenocorticotropic hormone (ACTH) stimulates the conversion of cholesterol to pregnenolone. The five types of steroid classes are progesterone, glucocorticoids, mineralocorticoids, androgens, and estrogens.

Cholesterol becomes pregnenolone, which then becomes progesterone. Progesterone can then either become testosterone, aldosterone, or cortisol. Testosterone can be further converted into estradiol. Progesterone is important for women to maintain pregnancy and the endometrial lining of the uterus, although it is also found in low levels in males. Cortisol (also known as hydrocortisone) is synthesized and secreted from cells in the cortex of the adrenal glands. In the liver, cortisol increases both glycogen synthesis and gluconeogenesis. In skeletal muscle, cortisol decreases glucose uptake and protein synthesis, and increases protein catabolism. In adipose tissue, cortisol increases lipid mobilization and decreases glucose uptake. Aldosterone is synthesized and released from cells in the adrenal cortex. It increases reabsorption of sodium ions in the kidneys, intestines, salivary glands, and sweat glands. This leads to retention of sodium in the extracellular fluid (ECF), thus increasing ECF volume, blood volume, blood pressure, and blood flow. Testosterone is important for sperm maturation in males and development of secondary sex characteristics. Estradiol is the primary estrogen in women and is important in the development of secondary sex characteristics and regulation of the ovarian cycle.

Cholesterol decreases membrane fluidity at high temperatures and increases membrane fluidity at low temperatures. Basically, the way to remember this is that whatever the thermodynamically expected behavior of the cell fluidity is at a given temperature, cholesterol decreases the magnitude of that effect so that the membrane behavior never gets too extreme.

In order to tolerate high temperatures, a thermophilic bacterium must have fatty acid tails that decrease the fluidity of its cell membrane. Saturated fatty acid tails have stronger Van der Waals interactions and thereby decrease the fluidity of cell membranes. Longer fatty acid tails have greater surface area for Van der Waals interactions, so they also decrease the

fluidity of cell membranes. Therefore, long saturated hydrocarbon tails would best help the thermophilic bacteria have stable plasma membranes.

Pinocytosis is the ingestion of liquid into a cell by the budding of small vesicles from the cell membrane.

5.6 Cellular Reproduction and Chromosomes—Eukaryotic Cells

Chromatin is a complex of dsDNA and histones (which are a type of protein). Histones are basic proteins consisting of a high percentage of Lys and Arg residues that bear a positive charge. This allows them to easily establish a favorable electrostatic relationship with the negatively charged DNA polymers. There are also nonhistone proteins (i.e. RNA polymerase) that are acidic and bear a net negative charge. Chromatin condenses into **chromosomes** as the cell prepares for division. At metaphase, each chromosome consists of two **sister chromatids**, each held together through a constricted region called the **centromere**.

There are five types of histones: H1, H2A, H2B, H3, and H4. A nucleosome is an association of these histones with a specific length of DNA. H2A, H2B, H3, and H4 are the **core histones**. The H1 histone is thought to play an important role in pulling the individual nucleosomes together to help with DNA condensation. Mitosis should honestly be pretty

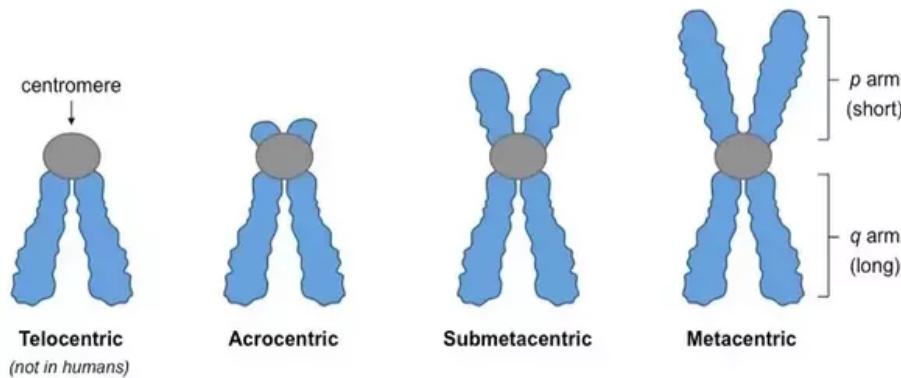


Figure 86: General classification of eukaryotic chromosomes based on the position of the centromere.

straightforward. The only thing is that for prophase, there may be some new terminology. At the beginning of prophase, the two **centriole** pairs begin to move apart, and microtubules begin to radiate from each pair in all directions, forming a star-like structure called an **aster**. The region from which the microtubules extend outward is called the **centrosome**, or MTOC (microtubule-organizing center). The microtubules ultimately form the **mitotic spindle**, and attach to the chromosomes at the **kinetochore**, a specialized area closely associated with the centromere.

5.7 Meiosis

Meiosis is technically composed of two processes—Meiosis I and Meiosis II—each of which is preceded by an interphase. During the second interphase immediately before meiosis II, the S period does not exist and so the DNA cannot be replicated again. Meiosis is a bit more complicated than mitosis, so let's take a look at it step by step:

- Prophase I is longer and more complicated than prophase of mitosis, and can be divided into five stages: leptotene, zygotene, pachytene, diplotene, and diakinesis. During *leptotene*, the replicated chromosomes have already started to condense and now become visible. During *zygotene*, the synaptonemal complex forms where the maternal and paternal chromosomes come together to form the tetrads. This is in preparation for crossing over, although crossing over has not occurred yet. At this stage, there are 23 tetrads, 2 centromeres in each tetrad, and 46 chromosomes (based on the number of centromeres). During *pachytene*, chromosomes continue to condense, and crossing over occurs. During *diplotene*, the homologous chromosomes begin to separate and the events of crossing over become visible at structures called **chiasmata**. During *diakinesis*, the nuclear envelope begins to break down, and the nucleoli disappear.
- Metaphase I is unremarkable.

- During Anaphase I, the microtubules pull the homologous chromosomes apart, but the centromeres do *not* divide. Each chromosome that migrates toward a pole is still composed of sister chromatids, and this pair is referred to as a **dyad**. Cytokinesis also begins at this step.
- Telophase I and Cytokinesis are unremarkable. After the completion of Meiosis I, we find that a diploid cell with 46 chromosomes has divided into two haploid cells, each with 23 chromosomes. This is therefore a **reductive division**.
- Interphase II is usually pretty brief.
- Meiosis II is pretty unremarkable. One thing to note is that the cells entering Meiosis II are already haploids. Also, because of the crossing over events from Meiosis I, the chromatids of each chromosome that are separating during Anaphase II can't actually be referred to as *sister* chromatids because their DNA is no longer the same because of the genetic recombination.

Yeast are an example of a eukaryote that reproduces asexually.

5.8 Biomembranes and Membrane Transport

In lipid bilayers, recall that lateral diffusion of neighboring phospholipids is very common, but transverse diffusion (flip-flop) of phospholipids from one lipid plane to the next is a very rare event. Addition of cholesterol to a membrane acts to decrease fluidity, as the planar steroid ring inserts between neighboring fatty acid side chains and interferes with the movement of those chains. It might be a good idea to review that one problem about the symports and antiports from the Bi9 Final.

There's also something called **group translocation** found in certain bacteria, where a sugar residue like glucose is phosphorylated as it is being transported through the plasma membrane. This type of transport is coupled to cellular metabolism. There's also **bulk transport** that involves *endosomes* and *endocytotic vesicles*. Many animal cells will show an invagination of a portion of their plasma membrane that will eventually pinch off to form an internalized vesicle. Bulk transport is where words like endocytosis, pinocytosis, phagocytosis, and exocytosis come in.

5.9 Nucleus, Nucleolus, & Ribosomes

The nucleus is double membrane-bound. The outer membrane is actually considered to become part of the rough endoplasmic reticulum (RER). Between the inner and outer membrane is the **perinuclear space**. DNA replication and transcription occur in the nucleus, while translation happens outside in the cytosol. The diameter of **nuclear pores** is 10-20 nm.

Within the nucleus is a highly organized region called the **nucleolus**. The nucleolus is *not* a membrane-bound organelle. It is centered around certain chromosomes that contain *nucleolus organizer regions*, and is involved in the synthesis of rRNA. If a cell is quite actively involved in protein synthesis (meaning that it needs a lot of rRNA), one would expect the nucleolus to be larger than if a cell were not as actively involved in protein synthesis.

Eukaryotic ribosomes are composed of two subunits, each differing in size and content of RNA and protein. We talk about ribosome size based on a sedimentation coefficient called the **Svedberg unit (S)**, where one S = 10^{-13} sec. The rate at which a molecule sediments in an ultracentrifuge tells us something about its mass. The sedimentation coefficient of a complete eukaryotic ribosome is expressed as **80S**. Eukaryotic ribosomes have two subunits—a large subunit (60S) and a small subunit (40S). Note that $60S + 40S \neq 100S$; in other words, the values are not additive. Sedimentation coefficients are not linearly related to molecular weight, as they depend on the size and the shape of the molecule.

The overall dimensions of a complete ribosome is about **20 nm × 30 nm**, and contains roughly 60% rRNA and 40% protein. The small subunit is about 9 nm in diameter and contains roughly half rRNA and half protein. The large subunit is roughly 25 nm in diameter and contains about 65% rRNA and 35% protein.

Ribosomes are also found in the matrix of the mitochondria. Ribosomes found in the mitochondrial matrix differ in RNA and protein content, and are also smaller and sediment at about **55S**. Prokaryotic ribosomes are about **70S**.

5.10 Endoplasmic Reticulum, Golgi Apparatus, & Lysosomes

The ER is the largest membrane system in a eukaryotic cell. The **Smooth Endoplasmic Reticulum (SER)** is tubular in shape and is involved in the synthesis of a majority of the cell's membrane lipids (i.e. neutral fats, phospholipids, prostaglandins, and steroid hormones). Especially in hepatocytes (liver cells), the SER is involved in hydroxylation reactions that aid in the detoxification of drugs by making them more water soluble. The SER can also help catabolize liver glycogen in hepatocytes.

The **Rough Endoplasmic Reticulum (RER)** is flat and sheet-like, and ribosomes on the cytoplasmic face of the RER are bound to the membrane by their large (60S) subunit. Post-translational modification of proteins occur in the ER lumen, where certain amino acids are modified by hydroxylation and glycosylation events.

For the Golgi Apparatus, the **cis** cisterna face the nucleus/ER, while the **trans** cisterna face the plasma membrane. The **medial** cisterna are located between the cis and trans cisternae. After a protein enters the cis cisterna and before it leaves the trans cisterna, it can be modified in a variety of different ways, including glycosylation, sulfation, and proteolysis.

Lysosomes contain special enzymes that function at acidic pH values, and are referred to as **acid hydrolases**, which catalyze the general reaction $A - B + H_2O \longrightarrow A - H + B - OH$. There are some common acid hydrolases that you should know: Lysosomal enzymes are inactive at neutral pH. The hydrolysis products simply diffuse out of the organelle and are

| Enzyme | Substrate | Bond Hydrolyzed | Example |
|--------------|-------------------|----------------------|-------------------------|
| Proteases | Peptides | Peptide | Pepsidase |
| Glycosidase | Glycolipids | Glycoside | β -hexosaminidase |
| Lipase | Phospholipids | Carboxylic Ester | Phospholipase |
| Nuclease | DNA | Phosphoric Diester | Acid Deoxyribonuclease |
| Phosphatases | Phosphomonoesters | Phosphoric Monoester | Acid Phosphatase |

utilized in a variety of metabolic processes. We know that if this were not the case, the increasing solute concentration in the lysosome would eventually result in the lysis of the lysosome due to osmotic water entry into the lysosome, but this is not observed.

Peroxisomes are membrane-bound organelles that have a number of enzymes, the most notable being *catalase* that catalyzes the degradation of hydrogen peroxide H_2O_2 .

5.11 Signal Hypothesis

The Signal Hypothesis was covered in Bi9, and is import for co-translational transport. Many proteins have a signal sequence that binds to a signal recognition particle (SRP) in the cytoplasm, halting translation. The SRP chaperones the complex to the signal sequence receptor embedded in the membrane of the RER, and the signal sequence is cleaved by signal peptidase. Translation then continues and at the same time, the polypeptide chain is transported into the ER lumen via Sec61.

5.12 Cytoskeleton

Microtubules are composed of 13 protofilaments and are about 25 nm in diameter. Each protofilament has alternating α -tubulin and β -tubulin proteins. The growth of microtubules occurs from **microtubule organizing centers (MTOCs)**. Three common centers are the centrosome (cell center), kinetochores (spindle attachment sites on chromosomes), and centrioles. Polymerization preferentially occurs at the (+) end.

Microfilaments (actin filaments) are about 7 nm in diameter. Polymerization preferentially occurs at the (+) end. We have G-actin (globular) monomers, and F-actin (filamentous) polymerized subunits.

Intermediate filaments are about 8 to 12 nm in diameter and differ in composition. For example, the intermediate filaments in epithelial cells are composed of keratins, while in muscle cells they are composed of desmin.

5.13 Prokaryotic Cells

The two most frequently encountered bacteria are the **cocci** and the **rods**. Cocci are essentially spherical in shape while rods generally resemble the shape of a tube. Bacteria that have a rigid twist to their rod-like structure are called **spirilla**. If their twisted structure is more flexible, they are called **spirochetes**. Invaginations of bacterial cell membranes are called **mesosomes** with currently unknown function.

Instead of membrane-bound organelles, prokaryotes often have structures called **inclusion bodies**, which can contain organic molecules like glycogen or inorganic molecules like phosphate granules. Prokaryotic ribosomes are **70S**, and are composed of a large **50S** subunit and a small **30S** subunit.

Gram positive bacteria have a rather thick, homogeneous **peptidoglycan** layer (20 nm to 80 nm) just outside their plasma membrane. **Gram negative** bacteria have a much thinner peptidoglycan layer (1 nm to 3 nm), in addition to an outer membrane that contains **lipopolysaccharides** and **porins**. The polysaccharide helps to stabilize the membrane, and also acts as an *endotoxin* and provides a defense mechanism for the cell.

Genetic material can be passed from one bacterial cell to the next by binary fission, bacterial conjugation, transformation, or transduction. Binary fission is basically asexual reproduction where one cell splits into two. In bacterial conjugation, genetic information is transferred by cell-cell contact. Donor strains of bacteria are F^+ (male), while the recipient bacteria are F^- (female). The "F" refers to the **fertility plasmid**. Transformation is the uptake of genetic material from the surrounding medium. Transduction is the transfer of bacterial genes by viruses.

5.14 Viruses

The architecture of a virus is usually based on one of two structural motifs: **isometric** (usually in the form of an icosahedron) and **helical**. The viral protein coat is formed from **capsomers**. If there is no nucleic acid within the protein coat/shell, then the empty shell is referred to as a **capsid**. However, if there is nucleic acid within the protein shell, the complex is called a **nucleocapsid**.

The genetic information within the genome of a virus may be encoded in either the language of DNA or RNA, and can be linear or circular, single- or double-stranded, and even segmented. However, no matter how the genetic information is stored in the virus, the translational process uses mRNA as a template. Therefore, by convention, we define that mRNA as being a **positive (+) strand** nucleic acid.

Naked (non-enveloped) viruses gain access to the host's cytoplasm only by receptor-mediated endocytosis. The receptors on the cell surface of the host are usually located near specialized depressions called **clathrin-coated pits**. Enveloped viruses can either enter through receptor-mediated endocytosis or by direct fusion with the plasma membrane. Release of viruses (whether naked or not) is mediated by pH changes, as the pH becomes more acidic in the endosome.

5.15 Metabolic Components—Enzyme Kinetics

The **Michaelis-Menten Equation** describes the rate of a reaction as a function of the substrate concentration, assuming that the concentration of the intermediates is in steady state (meaning that the rate of change of the intermediates is zero or negligible). The equation is

$$V = \frac{V_{max} [S]}{[S] + K_M} \quad (84)$$

where K_M is the Michaelis constant. Notice that when $[S] = K_M$, then $V = V_{max}/2$, so this means that K_M is equal to the substrate concentration at which the reaction rate is half of its maximal value. In general, a *high* K_M value indicates a *weak* binding of the enzyme-substrate complex, while a *low* K_M value indicates a *strong* binding of the enzyme-substrate complex.

The **Lineweaver-Burk Plot** is basically the reciprocal of the Michaelis-Menten equation, and has the form of a straight line:

$$\frac{1}{V} = \frac{K_M}{V_{max}} \frac{1}{[S]} + \frac{1}{V_{max}} \quad (85)$$

When an enzyme is completely saturated with substrate, then the number of substrate molecules which are converted to product per unit time is referred to as the **turnover number**.

There are two major types of enzyme inhibition: reversible and irreversible inhibition. There are two types of reversible inhibition: competitive and non-competitive inhibition. Allosteric inhibition is a subset of non-competitive inhibition.

Competitive inhibitors decrease the rate of catalysis of the enzyme, and can be overcome at high concentrations of substrate. Initially, we have a lower V_{max} for a given substrate concentration, but as we approach an infinite substrate concentration, the concentration of the competitive inhibitor becomes negligible and the enzyme catalyzed reaction will approach the same V_{max} as before. When a competitive inhibitor is present, it changes the *apparent* K_M of the enzyme. The word 'apparent' is used because each enzyme has a characteristic K_M value for a given substrate. In other words, the K_M of a competitively inhibited enzyme *appears* to shift to a higher value, but if we add enough substrate, we will eventually reach the same V_{max} . On a Lineweaver-Burk plot, the y -intercept remains the same but the slope increases due to the competitive inhibitor.

Non-competitive inhibition *cannot* be overcome by increasing the substrate concentration. This means that we will have a **lower** apparent V_{max} for the reaction. However, the K_M will not change. On the Lineweaver-Burk plot, this looks like the x -intercept stays the same but the slope increases due to the non-competitive inhibitor, due to the decreased V_{max} .

5.16 Metabolic Components—Enzymes and Molecules

There are four classic enzymes in biochemistry:

1. **Lysozymes** hydrolyze a specific glycosidic bond in the polysaccharide component of certain bacterial cell walls.
2. **Ribonuclease** hydrolyzes the phosphodiester bonds in RNA polymers.
3. **Carboxypeptidase** is a Zn-containing digestive enzyme secreted by the exocrine cells of the pancreas, and hydrolyzes the carboxyl terminal peptide bond of protein polypeptide chains.
4. **Chymotrypsin** hydrolyzes either ester or peptide bonds. If a peptide is hydrolyzed, the products will be an amine and an acid. If an ester is hydrolyzed, the resulting products will be an alcohol and an acid.

It is important to note that in some instances, RNA can also act as a catalytic enzyme.

Many enzymes are initially synthesized in the form of a **proenzyme** or **zymogen**. Zymogens are *inactive precursors* of the active enzyme. In order for these enzymes to become active, they must first undergo some type of proteolytic activation (which is an irreversible step).

Transition state analogs are synthesized with the hopes that they will occupy the active site of an enzyme. The role of an enzyme is to bind the transition state and stabilize it. A transition state analog is thus a *competitive inhibitor* of the actual reaction.

Metabolism is the generalized word for all of the processes which occur inside living organisms. **ATP** is an important metabolic molecule composed of **adenine** (a base), **D-ribose** (a sugar), and three **phosphate groups**. Adenine and D-ribose taken together make up the molecule **adenosine**. The linkage between the C1' of the ribose ring and the N9 of the adenine ring is called an **N-acetal linkage** (this is likely not super important).

Enzymes often require substances other than amino acids to carry out their functions. These substances are called **cofactors**. For example, consider an enzyme with an active site—this active site is isolated from the rest of the enzyme, and within its cavity are *hydrophobic* side chains. If you have a substrate which is in solution, it tells you that it must have some degree of polarity; otherwise, it wouldn't be soluble in solution. Suppose you wanted to do something to the substrate that required a non-polar environment. One way to achieve that situation in a cell that is full of water is to get the substrate to settle down within the pocket of the active site (which is surrounded by hydrophobic residues). If there are things that you would like to do to the substrate that *cannot* be done with the canonical 20-amino-acid tool kit, this requires a **cofactor**. If these cofactors are required by an enzyme, they are called **coenzymes**. A coenzyme is a non-amino acid, non-polypeptide, non-protein substance which is *required* for the activity of an enzyme. In many cases, the coenzyme is bound to the enzyme in a non-covalent manner, but there are also many cases in which the coenzyme is covalently attached. There also may be more than one coenzyme involved with a single enzyme.

One coenzyme that is very important to us is **nicotinamide adenine dinucleotide (NAD)**. Humans have the ability to synthesize the nicotinamide (also called niacinamide) portion of the NAD molecule (the actual reactive part of the molecule), provided that they have enough of the amino acid *tryptophan* in their diet—for this reason, tryptophan is considered to be an **essential amino acid**. Once we have produced nicotinamide, the rest of the NAD molecule can be synthesized. If we don't have enough of the NAD coenzyme to carry out our metabolic reactions, then we will not function as well. This is the basis of a **vitamin deficiency**.

The coenzyme **NAD** is held at a specific site on the surface of the active site of an enzyme. NAD is held there by its **R group**, which is the adenine dinucleotide portion. You can kind of imagine NAD as a screwdriver, with the adenine dinucleotide being the head of the screwdriver and the niacinamide portion being the part that actually engages the screw. In this analogy, the enzyme is ‘holding’ the NAD coenzyme by its handle, and positions the nicotinamide portion toward the incoming substrate in the active site.

Another important metabolic molecule is **flavin adenine dinucleotide (FAD)**, which is a derivative of **riboflavin** (vitamin B₂). We cannot synthesize riboflavin ourselves, and instead must obtain it elsewhere (e.g. in the diet). Riboflavin becomes a component in the structure of a molecule called **flavin mononucleotide (FMN)**, which is basically just phosphorylated FMN. FMN is an active coenzyme. If we were to add another phosphate, a ribose ring, and an adenine base to FMN, we would get the coenzyme **flavin adenine dinucleotide (FAD)**.

Another coenzyme that is important to us is **Coenzyme A (CoA)**. The terminal sulphydryl group is the active site of CoA, and the rest of the molecule can be thought of as being like a handle, similar to the handle on a screwdriver. *Acyl groups* can be linked to the sulphydryl function to form acyl CoA. A common acyl group that is linked to CoA is an acetyl moiety, thus giving **acetyl CoA**. The structure of CoA is shown in Fig. 5.16.

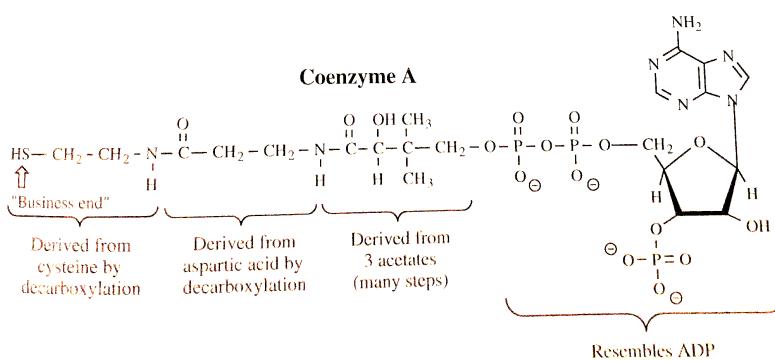


Figure 87: The structure of coenzyme A.

5.17 Metabolic Pathways

Glycolysis: Glycolysis occurs in the cytosol and can be divided into two phases, **Phase I** and **Phase II**. **Phase I** comprises steps I through V and **Phase II** comprises steps VI through X. Note that Phase 1 does *not* change the oxidation states of the carbons—there are no oxidation or reduction reactions involved.

1. The enzyme **hexokinase** puts a phosphate group from ATP onto the C-6 carbon of D-glucose, giving us **glucose-6-phosphate (G6P)**. This uses 1 ATP and is therefore known as an *investment step*. Once this step occurs, the glucose is committed to glycolysis. This is because glucose does not have any charges and can pass back and forth across the cell's membrane relatively easily. However, once it is phosphorylated, it picks up some negative charges and can no longer pass across the cell's membrane. Thus, glucose (in the form of G6P) is trapped inside the cell. This step is thermodynamically ***favorable***.
2. Equilibrium isomerization between glucose-6-phosphate (G6P) and fructose-6-phosphate. The reaction takes place at the C-1 and C-2 carbons of G6P and proceeds through an enediol intermediate. The compound that emerges from this reaction is **fructose-6-phosphate**, a ketose (G6P is an aldose). The enzyme that catalyzes this is **phosphoglucone isomerase**. This step is thermodynamically ***unfavorable***.
3. The enzyme **phosphofructokinase** (part of the general class of **transferases**) catalyzes the conversion of fructose-6-phosphate into **fructose-1,6-diphosphate**. Basically, another phosphate group is put on, representing a *second investment* of ATP and is another *control point* in glycolysis. This step is thermodynamically ***favorable***.
4. The bond between C-3 and C-4 of fructose-1,6-diphosphate is broken through a **reverse aldol condensation**, giving **dihydroxyacetonephosphate (DHAP)** and **glyceraldehyde-3-phosphate (G3P)**. Both of these are *triooses*. G3P is an aldose and DHAP is a ketose. The enzyme involved here is **aldolase**, and is part of the general class of **lyases**. This step is thermodynamically ***unfavorable***.
5. DHAP is converted into G3P, as these carbohydrates are isomers. This reaction is catalyzed by **triose phosphate isomerase**. This isomerization to G3P is thermodynamically ***unfavorable***. All of the other significant reactions in glycolysis proceed from G3P. Even though about 96% of the triose phosphate molecule at equilibrium is DHAP, once G3P is removed by subsequent reactions, DHAP is ‘pushed’ to isomerize to G3P by Le Châtelier’s principle. Thus, from one molecule of fructose-1,6-diphosphate, we can get two molecules of G3P.
6. C-1 carbon of G3P is oxidized. The enzyme **glyceraldehyde-3-dehydrogenase** converts G3P into **1,3-diphosphoglycerate (1,3-DPG)**. Note that 1,3-DPG has a *mixed anhydride linkage* and a *monophosphoester linkage*. The additional phosphate group that gets tacked on comes from *inorganic phosphate Pi*, not from ATP. The oxidizing agent in this reaction is NAD^+ . This reaction is thermodynamically ***unfavorable***. This reaction is shown in Fig. 5.17.
7. The mixed anhydride linkage in 1,3-DPG is relatively unstable, so the hydrolysis of 1,3-DPG is *highly* thermodynamically ***favorable***. The enzyme involved is **phosphoglycerate kinase**, and gives the Pi to ADP to give ATP. This gives us 2 ATPs (one from each G3P molecule from before step 6).

8. 3-phosphoglycerate is converted to **2-phosphoglycerate** by the enzyme **phosphoglyceromutase**. Note that the phosphate group in the product is *not* the same phosphate group from 3-phosphoglycerate—this is because the enzyme itself is a *phosphorylated enzyme*, and during an intermediate state, the phosphorous from the enzyme gets transferred to the C-2 position of 2-phosphoglycerate. This reaction is thermodynamically ***unfavorable***.
9. Loss of water from 2-phosphoglycerate. This involves the hydroxyl function at the C-3 carbon and the hydrogen at the C-2 position. The enzyme **enolase** catalyzes the reaction to give **phosphoenolpyruvate (PEP)**. PEP has an *unstable, high energy* phosphate bond due to pronounced electrostatic repulsion. This reaction is thermodynamically ***unfavorable***. The equilibrium between the enol and keto forms of PEP very much favors the keto form.
10. The enzyme **pyruvate kinase** catalyzes the transfer of a phosphoryl group from PEP to ADP to give us **pyruvate** and **ATP**. This phosphorylation is ***nonoxidative***. This reaction is overall thermodynamically ***favorable***. This step also gives us another 2 ATPs (one from each G3P molecule from before step 6).

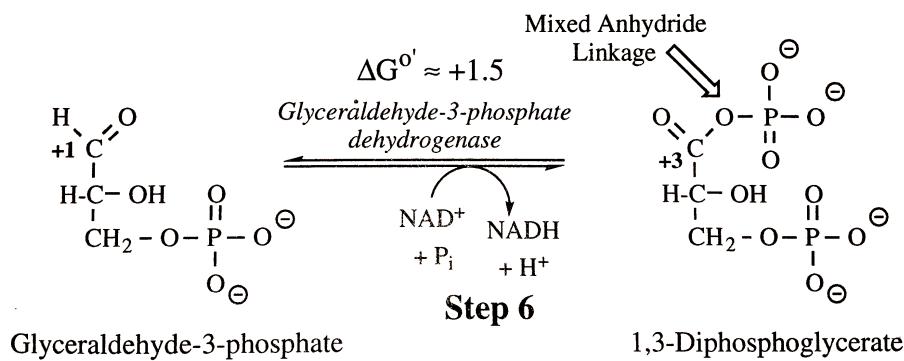


Figure 88: The sixth step of glycolysis. G3P is oxidized to 1,3-DPG. Notice the mixed anhydride linkage and monophosphate linkage.

After this process, we consumed 2 ATPs (1 in step 1 and 1 in step 3), and got back 4 ATPs from our initial ‘investment’ (2 in step 6 and 2 in step 10). **Our net profit is thus 2 ATPs.** In step 6, we also used 2 molecules of NAD^+ to give two molecules of NADH and H^+ . In order to recycle NADH back into NAD^+ to continue glycolysis, we must have some way to restore the reduced NAD’s back to their oxidized form. There are several ways to do this, and pyruvate is involved in all of them.

If we use pyruvate itself to regenerate the NAD^+ , then pyruvate will get *reduced*. Look at the C-2 carbon of pyruvate. It is in the keto form, and the oxidation state in this form is +2. If we did a **2 electron transfer**, we would end up with a carbon that has a zero oxidation state. The compound that we can produce from this reaction is **L-lactate**, and the enzyme involved here is **lactate dehydrogenase**. This is one of the possible fates of pyruvate. In this way, we can reoxidize NAD and pay off our debt from step 6. This reaction occurs in cells when O_2 is a *limiting factor*.

Because lactate cannot be reduced any further, it is a dead end in metabolism, and as we will later discover, is transported by the blood to the liver where it is converted back to pyruvate. Pyruvate will then be converted into glucose by a process called gluconeogenesis, which we will learn about later.

This is not the only way to regenerate NAD^+ . Yeasts use a slightly modified procedure. Instead of doing a direct reduction of pyruvic acid with reduced NADH, yeasts first *decarboxylate* the pyruvic acid to **acetaldehyde**—this reaction is catalyzed by **pyruvate decarboxylase**, a type of **lyase**. Acetaldehyde will next react with the reduced NADH to form **ethanol**. The enzyme involved in this step is **alcohol dehydrogenase**. This process is called *alcoholic fermentation*, and ethanol, a waste product, is excreted into the surrounding medium.

There is a certain level of control of the glycolytic pathway. High levels of ATP tend to ***allosterically inhibit*** phosphofructokinase in step 3. High levels of H^+ also inhibit this enzyme as well. One place we could get an increase in hydrogen ions is from the conversion of pyruvate to lactate. If you were to produce too many hydrogen ions, blood pH would begin to drop and you would experience **acidosis**. One other regulatory molecule of phosphofructokinase is **citrate**, an intermediate in the **Krebs cycle** that we’ll talk about later. If the levels of citrate are high, it must mean that glycolysis is functioning at some optimal rate because there is plenty of glucose around, and why waste the glucose that has been made available? Thus, high levels of citrate also inhibit phosphofructokinase. Conversely, if the levels of ATP are low, then it is often converted to

ADP and even AMP if absolutely necessary. Thus, it turns out that high levels of AMP stimulate phosphofructokinase.

Glycolysis is a long and complicated metabolic process. The overall picture of glycolysis is shown in **Fig. 5.17**.

Disaccharide Metabolism: Disaccharides include **maltose**, **sucrose**, and **lactose**, and can be hydrolyzed into

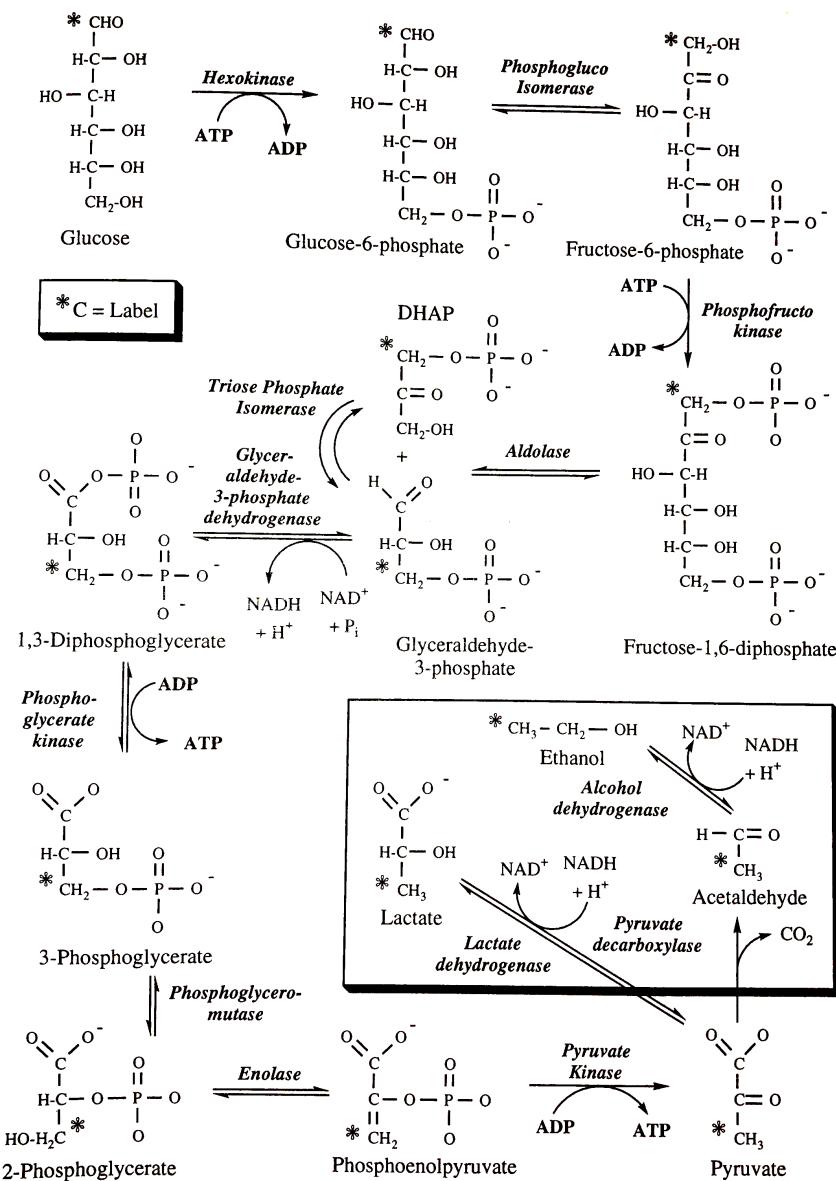


Figure 89: Overview of glycolysis. The labeled carbon is a radioactive C-1 to help us follow a particular carbon throughout the pathway.

their constituent monosaccharide residues which can then enter into the glycolytic pathway. Since maltose is made up of two β -D-glucopyranose molecules, these residues are immediately ready to enter into the glycolytic pathway. However, the β -D-fructofuranose residue from sucrose and the β -D-galactopyranose residue from galactose both need to be converted to a form which can enter the glycolytic pathway. Fructofuranose from sucrose can be converted into either (1) β -D-fructofuranose-6-phosphate by the enzyme **hexokinase**, or (2) β -D-fructofuranose-1-phosphate by the enzyme **fructokinase**. Both reactions are phosphorylation reactions and both require ATP. These can then split into DHAP and D-glyceraldehyde, which end up as G3P that can then enter the glycolysis pathway. For the galactopyranose residue from lactose, it can be converted to glucose-6-phosphate (G6P) in four steps, since galactose is an *epimer* of glucose at the C-4 position. G6P is the natural committed starting point for glycolysis, and so the converted galactopyranose can now enter the glycolysis pathway.

Krebs Cycle: The **citric acid cycle** (also the **tricarboxylic acid cycle** or the **Krebs cycle**) occurs inside the **mitochondria**. The first reaction that we need to look at is the *oxidative decarboxylation* of pyruvate to form **acetyl CoA**. The process occurs stepwise:

1. Thiamine pyrophosphate (**TPP**) adds to the carbonyl function of pyruvate.
2. The resulting addition compound **decarboxylates** and the remaining portion of pyruvate (a hydroxy-ethyl moiety) is transferred to **lipoamide**, where it is oxidized to form acetyl-lipoamide.
3. The acetyl moiety of acetyl-lipoamide is next transferred to CoA to form **acetyl-CoA** and **dihydrolipoamide** (the reduced form of lipoamide). NAD⁺ is converted to NADH + H⁺ as it reoxidizes dihydrolipoamide to lipoamide.

The important parts of this reaction can be seen in **Fig. 5.17**. This reaction is thermodynamically favorable.

The two-carbon acetyl CoA is unique because it has a **high energy thioester bond**. We want to basically use the

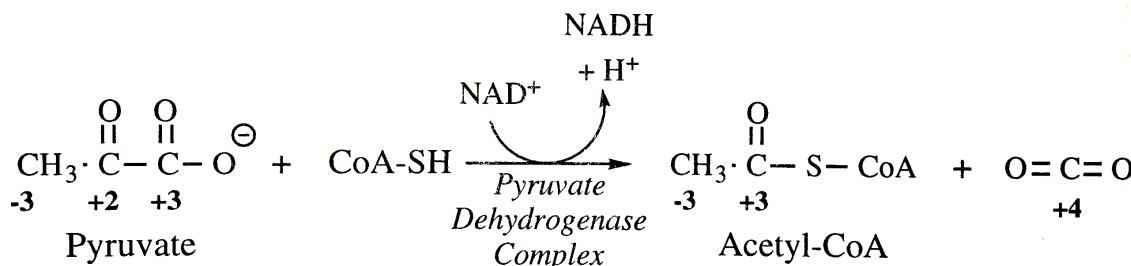


Figure 90: Overview of conversion of pyruvate to acetyl CoA. This oxidation reaction is coupled to the reduction of NAD⁺ to NADH.

energy stored in this bond to turn the two carbons into CO₂. The only way to accomplish this is to extend the carbon chain, which is done by combining acetyl CoA with a four-carbon acceptor molecule called **oxaloacetic acid (OAA)**. The α hydrogen of acetyl-CoA will form a bond with the carbonyl carbon of OAA, as shown in **Fig. 5.17**. The reaction proceeds via an **aldol condensation** to form the intermediate **citryl-CoA**, which is then hydrolyzed to form **citrate** with the subsequent regeneration of CoA. **It is the hydrolysis of the thioester bond that allows this reaction to be favorably pulled to completion.** The enzyme that catalyzes this reaction is **citrate synthetase**—it is in the class of enzymes called **lyases**.

Notice that citrate is a symmetrical molecule, but even so, the two $-\text{CH}_2-\text{COO}^-$ groups do not react identically, due

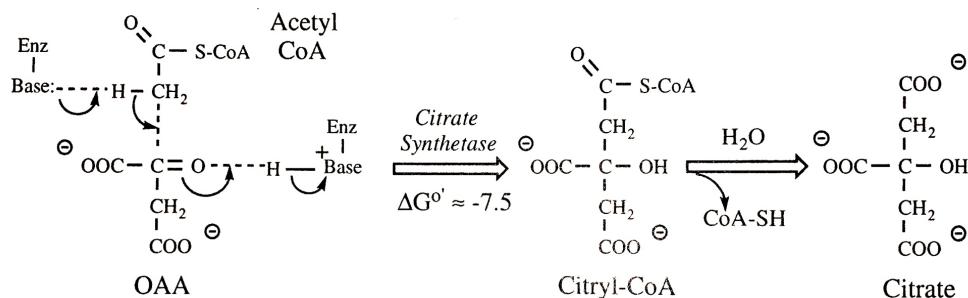


Figure 91: Formation of citrate from reaction of oxaloacetic acid (OAA) and acetyl CoA.

to the asymmetric nature of the enzyme **aconitase** in relationship to the symmetrical nature of citrate. Aconitase is the enzyme that converts citrate to **cis-aconitate**, and can distinguish between the two symmetrical groups in citrate. Thus, we can think of citrate as a *prochiral* molecule and it is optically inactive. Cis-aconitate is formed in the reaction via a E₁ elimination reaction. Notice that we can then add water across the double bond in cis-aconitate to form **isocitrate**, a process that is also catalyzed by the enzyme aconitase. Isocitrate has a secondary hydroxyl group that is capable of being oxidized to a carbonyl function. It is oxidized with NAD⁺ into **β -keto acid** by **isocitrate dehydrogenase**. Because the β -keto acid is pretty unstable, it will result in the spontaneous *decarboxylation* at the C-3 position of isocitrate to give **α -ketoglutarate (α -KG)**. This whole process is shown in **Fig. 5.17**.

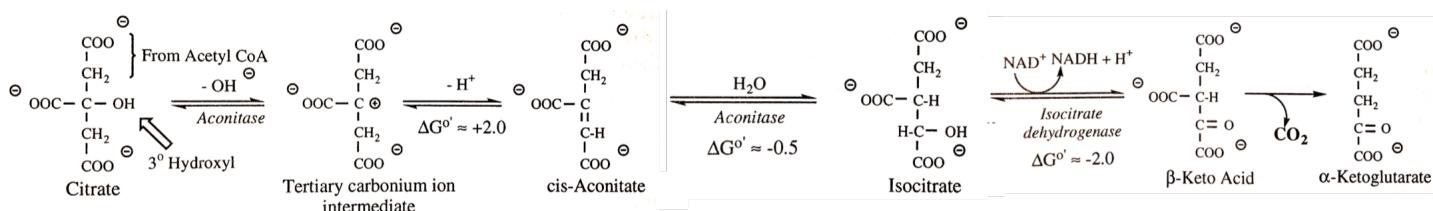


Figure 92: Citrate reacts with aconitase enzyme to give cis-Aconitate, and water is added across its double bond to give isocitrate. Isocitrate is then oxidized to give β -keto acid, and spontaneous decarboxylation gives α -ketoglutarate.

We now have an α -keto acid, which oxidizes to get **succinyl-CoA** and CO_2 . This reaction is catalyzed by the **α -ketoglutarate dehydrogenase complex**. Even though we are forming a high-energy thioester bond, this reaction is driven forward thermodynamically by the release of gaseous CO_2 , which is highly favored entropically. We can then take advantage of the resulting high-energy thioester bond in succinyl-CoA to make some **GTP**, which can readily be converted into **ATP**. The enzyme that catalyzes this conversion of succinyl-CoA to **succinate** is **succinyl-CoA synthetase**.

At this point, both of the carbons from acetyl CoA that we started with have already been released as 2 CO_2 molecules, so now we need to reconstruct succinate back into OAA so we can repeat the Krebs cycle. This occurs step-wise:

1. Oxidize succinate into **fumarate** the enzyme **succinate dehydrogenase**. Our oxidizing agent for this dehydrogenation reaction is **flavin adenine dinucleotide (FAD)**.
2. Fumarate reacts with water to produce **malate** in a thermodynamically favorable reaction. This is catalyzed by the **fumarase** enzyme.
3. We oxidize the hydroxyl group on malate into a ketone by coupling it to the reduction of NAD^+ . This is the reaction that will return us to OAA and bring us full cycle. The enzyme involved here is **malate dehydrogenase**. OAA is now ready to react with another acetyl-CoA molecule.

At this point, all of our energy is stored in the $FADH_2$ and $NADH$ reduced coenzymes. Even though we have made only 2 GTP's by going through the Krebs cycle, we have stored a great deal of potential energy. We have summarized the essentials of glycolysis and the Krebs cycle in **Fig. 5.17**.

Electron Transport: The outer membrane of mitochondria is quite permeable, and it contains protein channels called **porins** that allow small molecules to pass through. Molecules such as amino acids, carbohydrates, and small polypeptides can pass through these porins. The intermembrane space has a variety of enzymes and also a *high proton concentration* because of the protons that are pumped into this space from the matrix. The matrix of mitochondria contains hundreds of different types of enzymes, and is a region rich in enzymatic activity in general. The Krebs cycle that we just discussed occurs in the matrix.

Recall that in glycolysis and the Krebs cycle, glucose was essentially oxidized to give either lactate or CO_2 , but the important thing is that it was oxidized. In order to balance this, the electrons helped reduce either $NADH$ or $FADH_2$. In order to recycle these electron carriers for future use, we must reoxidize $NADH$ and $FADH_2$ back to NAD^+ and FAD, respectively. The external oxidizing agent that we will use for this reoxidation is **oxygen**, thus reducing O_2 into the H_2O water byproduct in aerobic respiration. This process, which is referred to as **electron transport** and **oxidative phosphorylation**, occurs on the **inner membrane** of the mitochondria. It is worth noting that because prokaryotic cells do not have mitochondria organelles, they carry out the reactions of electron transport and oxidative phosphorylation on their cytoplasmic membrane.

Associated with the inner mitochondrial membrane are sequences of protein complexes that act as electron carriers. Let's consider each of these enzyme complexes:

1. The **NADH-Q reductase complex (Complex I)** take the electrons from reduced $NADH$ two pass two electrons and also at the same time two hydrogens to the oxidized **flavin mononucleotide (FMN)** prosthetic group (i.e. electron carrier) associated with this enzyme complex. This results in the production of reduced **$FMNH_2$** and the regeneration of oxidized NAD^+ .
2. The electrons from $FMNH_2$ are passed to a series of **iron-sulfur clusters** in which the Fe atoms cycle between the reduced ferrous and oxidized ferric states. The electrons will eventually be passed from a reduced Fe-S moiety to the

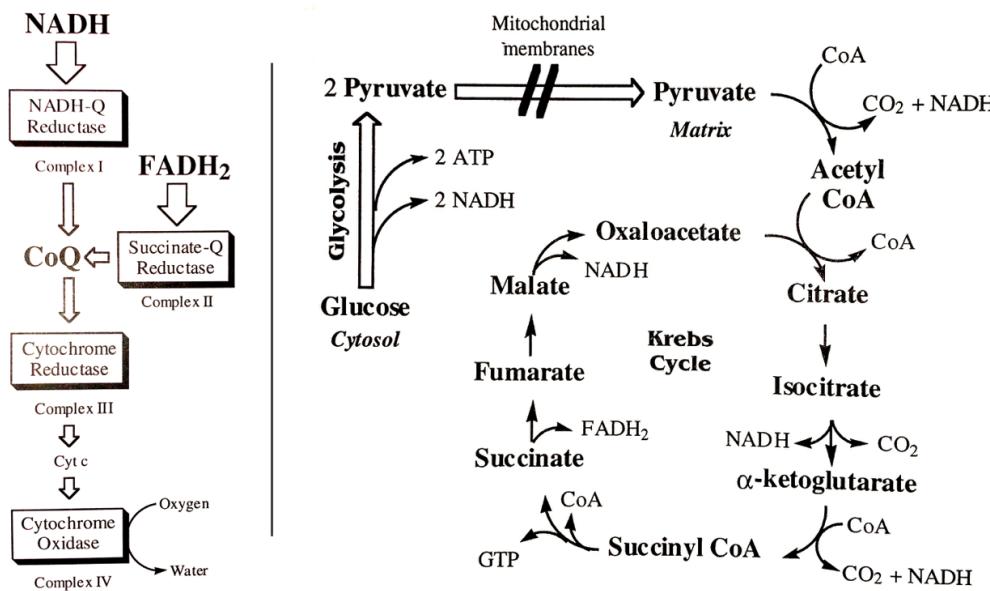


Figure 93: Left Overview of electron transport chain. Right Overview of the Krebs Cycle.

oxidized form of **coenzyme Q** (CoQ), a **quinone**. CoQ is also called **ubiquinone**, while CoQ in the reduced form (CoQH₂) is called **ubiquinol**.

- Similar to NADH, FADH₂ also needs to be reoxidized, and is done so by way of the **succinate-Q reductase complex (Complex II)**. The enzyme is **succinate dehydrogenase** catalyzes the transfer of electrons from FADH₂ to an Fe-S protein(s) which then funnels them into the oxidized form of CoQ. In the process, FADH₂ is reoxidized to FAD.
- What the first three steps we listed do is essentially collect all of the electrons from all of the various substrates into a *single* pocket, which is **CoQH₂**. Our next step, then, is to reoxidize this species back into CoQ. From here on out, the electron carriers between CoQ and oxygen are called **cytochromes**, which are electron transporting proteins that contain a *heme prosthetic group* with an iron atom that again alternates between the ferrous and ferric oxidation states.
- Within the **cytochrome reductase complex (Complex III)** are two cytochromes—**b** and **c₁**—and an Fe-S protein. As CoQH₂ transfers one electron at a time to a shuffling Fe-S protein in the complex, it is converted to CoQH[•] (a **semiquinone**). The reduced form of **cytochrome b** reacts with the semiquinone to give the oxidized form of cytochrome b and CoQH₂. This oxidized cytochrome b can then oxidize another molecule of CoQH[•] to CoQ. The electrons are eventually passed to **cytochrome c**.
- The electrons that end up on the reduced form of cytochrome c are next passed to the **cytochrome oxidase complex (Complex IV)**. This complex consists of two heme groups and two copper atoms—one associated with each heme group. Thus, the electrons are passed from cytochrome c to the heme moieties of the cytochrome oxidase complex, and then finally to **oxygen**. The transfer of four electrons to molecular oxygen leads to its reduction to two molecules of water. At this point, we have completed electron transport and have reoxidized the reduced coenzymes.

An overview of the electron transport chain is shown in Fig. 5.17.

Oxidative Phosphorylation: The passage of two electrons from NADH down the electron transport chain allows for the synthesis of 3 molecules of ATP, and the passage of two electrons from FADH₂ allows for the synthesis of 2 molecules of ATP. All of the other ATPs that we have made to this point from glycolysis and the Krebs cycle are referred to as **substrate level phosphorylations**—technically they don't directly depend on oxygen.

ATP synthesis and the electron transport chain are coupled together by a **protein gradient** that had established itself across the inner mitochondrial membrane and not by a high energy phosphorylated intermediate. The coupling factor for this mechanism is an enzyme called **F₀F₁ ATPase**—this enzyme allows for the synthesis of ATP at the expense of the free energy that is released as a proton passes through this F₀F₁ ATPase from the intermembrane space to the matrix of the

mitochondrion.

If there are reduced substrates (e.g. NADH, FADH₂) contributing electrons either from within the mitochondrion or from outside (e.g. through glycolysis), then the electrons will move down the electron transport chain and eventually will be used to reduce water. In this process, there is a rearrangement of these enzymes as they pass back and forth through their oxidized and reduced states, such that hydrogen ions are removed from the matrix into the intermembrane space. Hydrogen ions are thought to be translocated across the inner mitochondrial membrane at each the three enzyme complexes **NADH-Q Reductase**, **Cytochrome Reductase**, and **Cytochrome Oxidase**.

By oxidative phosphorylation, the total amount of ATP we've produced from aerobic respiration is 38 ATPs per glucose molecule. The 2 NADHs from glycolysis gives us 6 ATPs, while the 8 NADHs from the Krebs cycle will give us 24 ATPs. The 2 FADH₂s from the Krebs cycle will give us 4 ATPs. We have 2 ATPs made in glycolysis, and we have 2 GTPs (which can be converted into ATP) from the Krebs cycle. Therefore, the total ATPs produced is 38.

Pentose Phosphate Pathway: The purpose of the pentose phosphate pathway is to generate reducing power in the form of **NADPH** and five carbon sugars such as ribose-5-phosphate. These reactions occur in the *cytosol* of the cell, and some of them will be involved in the formation of other sugar molecules from CO₂ in photosynthesis.

Recall the formation of G6P by phosphorylation of glucose by hexokinase. G6P can be used in the pentose phosphate pathway. If we *dehydrogenate* G6P at the C-1 carbon, we get a lactone called **6-phosphoglucono- δ -lactone**. This reaction also yields **NADPH** and is catalyzed by **glucose-6-phosphate dehydrogenase**.

The next reaction is the **hydrolysis** of the lactone, catalyzed by **lactonase**. The opens the ring and gives us **6-phosphogluconate**. Recall from organic chemistry that the best way to release CO₂ is to have a keto function at the β position. Thus, **6-phosphogluconate dehydrogenase** makes this **β -keto acid** intermediate, catalyzed by **NADP⁺**. The transition state is the β -keto acid, which rapidly decarboxylates to form **ribulose-5-phosphate**. This is an **oxidative** way to form 5-carbon sugars.

Phosphopentose isomerase then isomerizes ribulose-5-phosphate to **ribose-5-phosphate**, similar to the interconversion between G6P and fructose-6-phosphate in glycolysis. Once we have the ribose-5-phosphate, it can eventually be incorporated into biomolecules such as RNA, DNA, ATP, and NAD. The pentose phosphate pathway, which occurs in the *cytosol* of cells, is one way to get five-carbon sugars and **NADPH**.

It is worth noting that as oxygen becomes reduced, it goes through a number of intermediate stages such as the superoxide radical, H₂O₂, and hydroxyl free radical. These intermediate species are toxic and can cause cellular damage. Compounds like **vitamin C**, **vitamin E**, and **ergothioneine** can act as free radical scavengers/anti-oxidants. **Superoxide dismutase (SOD)** can then catalyze the conversion of the superoxide radical into hydrogen peroxide and oxygen. Hydrogen peroxide can react with the enzyme **catalase** to be converted into water and oxygen.

Glucconeogenesis: Glucconeogenesis is the synthesis of glucose from non-carbohydrate precursors, such as **lactate**, **amino acids** like alanine, and **glycerol**. In the liver, lactate is converted to pyruvate, which can be converted into **glucose** by way of gluconeogenesis. The series of associated reactions is referred to as the **Cori cycle**, and is shown in Fig. 5.17.

Molecules like lactate, alanine, and glycerol can be converted into glucose because they all contain three carbon atoms.

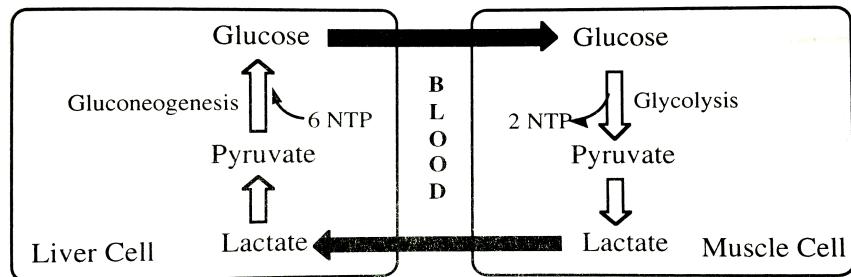


Figure 94: Overview of the Cori cycle.

It turns out that animals *cannot* get a net conversion of two carbon compounds (e.g. acetyl CoA) into glucose—they lack the necessary enzymes that will allow the conversion of a two-carbon unit into glucose. However, plants and many bacteria can make use of these two-carbon precursors because they have the enzymes necessary.

The first step in gluconeogenesis is accomplished by the enzyme **pyruvate carboxylase**, which is located inside the mitochondrial matrix of the liver cells. It contains a **biotin prosthetic group** that carries activated CO₂. Pyruvate then diffuses into the mitochondrial matrix and is *carboxylated* to form **OAA**. Pyruvate carboxylase is an example of a **ligase**

enzyme because a high energy bond is being used—if energy was not used, the enzyme would be called a *lyase*. This process is shown in **Fig. 5.17**. OAA can then enter the Krebs cycle and condense with acetyl CoA, leading to the eventual synthesis of more ATP. However, if the cell has *plenty* of ATP, then OAA will be utilized for gluconeogenesis.

In the case of gluconeogenesis, OAA is reduced to malate by NADH-linked **malate dehydrogenase**. Malate is trans-

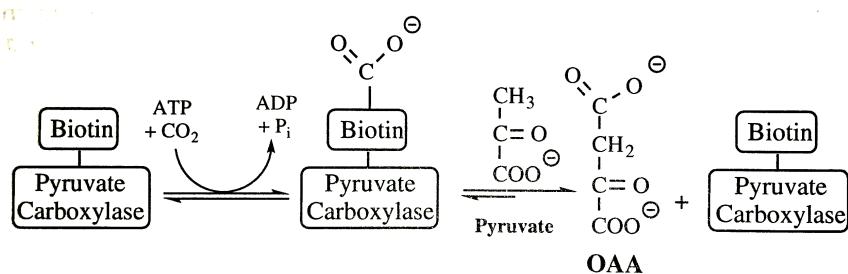


Figure 95: Converting pyruvate into OAA using pyruvate carboxylase.

ported back to the cytosol, and then reoxidized back into OAA by NAD⁺-linked malate dehydrogenase (note the change in the NAD⁺ oxidation). OAA is decarboxylated and phosphorylated by GTP to give **phosphoenolpyruvate (PEP)**. This reaction is catalyzed by **PEP carboxykinase**. Once we have this PEP, the rest of the reactions in glycolysis are reversible, and the equilibrium will favor moving back towards glucose until fructose-1,6-diphosphate is reached. This is true as long as there are high levels of ATP, as high levels of ATP turn out to *allosterically inhibit* **pyruvate kinase**, the enzyme that converts PEP to pyruvate.

From here, the fructose-1,6-diphosphate is hydrolyzed at the C-1 position to give fructose-6-phosphate, which is in equilibrium with glucose-6-phosphate (G6P). G6P can be used to make glycogen or it can be converted into glucose. The conversion of G6P into glucose is catalyzed by **glucose-6-phosphate phosphatase**. Overall, gluconeogenesis represents a *large potential futile loop* because we are losing NTPs in this process and could potentially get nowhere if we happen to get stuck in one of the steps. Overall, gluconeogenesis will cost us 6 ATPs, while glycolysis will give us a net yield of 2 ATPs. Therefore, it is really only favorable for gluconeogenesis to occur if times are good and ATP is plentiful.

Fatty Acid Oxidation: There is more energy available per unit weight of triglyceride than of hydrates like glycogen. If the amount of energy that had been concentrated in neutral fats where stored in the form of *glycogen*, you would be grossly overweight. This is because fats are at a lower oxidation state—the average oxidation state of carbons in a fatty acid molecule is -2, while in carbohydrates, it is 0. So, in the process of turning that carbon into CO₂, there is a lot more oxidation taking place for fats. The result is more energy being produced from the burning of a gram of fat compared to a gram of carbohydrate. Fats are also fairly hydrophobic, and so when they are stored there is not very much water. Fats also have to be metabolized aerobically, unlike carbohydrates.

The pathway in which fat is metabolized involves eight enzymatic steps:

1. Hydrolyze the carboxyester bonds using the enzyme **lipase** to separate glycerol from the three fatty acids. Both glycerol and fatty acids will eventually be converted into acetyl CoA, but by different series of reactions.
2. Glycerol is phosphorylated by **glycerol kinase** and ATP to produce glycerol-3-phosphate. The coenzyme NAD⁺ reacts with glycerol-3-phosphate to form **DHAP** (dihydroxyacetone phosphate) in the cytosol. DHAP can then go on to make pyruvate using the same reactions as in glycolysis.
3. The second thing we want to consider is the *activation* of fatty acids. We find that fatty acids are degraded 2 carbons at a time. The way that this is accomplished is that CoA forms a **thioester** bond with the carboxyl group of a fatty acid, driven by ATP hydrolysis into **AMP** (meaning two phosphate groups are lost). The activation of the fatty acid by bonding it to CoA is complete, and took place on the outer mitochondrial membrane. This is catalyzed by the enzyme **acyl CoA synthetase**.
4. Activated fatty acyl CoA are shuttled across the inner mitochondrial membrane by **carnitine**. Once the activated fatty acid is released in the mitochondrial matrix, carnitine will return to the cytosolic medium and the process will repeat itself.
5. Acyl-CoA is oxidized by **FAD** and **acyl-CoA dehydrogenase** to give **enoyl-CoA** and FADH₂. This process is pretty analogous to the conversion of succinic acid to fumaric acid in the Krebs cycle.

6. The double bond in enoyl-CoA is hydrated, catalyzed by **enoyl-CoA hydratase**. This gives us L-hydroxyacyl-CoA.
7. The hydroxyl group we added is oxidized to give a keto group, catalyzed by **L-3hydroxyacyl-CoA dehydrogenase**. The coenzyme involved is NAD^+ . This gives us **ketoacyl-CoA**. This reaction is analogous to the oxidation of malate to oxaloacetate in the Krebs cycle.
8. Ketoacyl-CoA reacts with a second CoA-SH molecule to cleave the C–C bond between the α and β carbons. This gives us an acetyl-CoA molecule (which can be utilized in the Krebs cycle) and a **fatty acyl-CoA** molecule that has been *shortened by 2 carbon atoms*. This *thiolytic cleavage* is catalyzed by **β -ketothiolase**. The fatty acyl-CoA molecule that we just produced is *already activated*, which means that we don't have to repeat the activation step again.

This process makes a lot of sense for fatty acids that have an even number of carbons, since they are removed two at a time. However, for fatty acids that have an odd number of carbons, we have a remaining three-carbon **propionyl-CoA** residue, and we need a way to metabolize it:

1. Carboxylate propionyl-CoA into **D-methylmalonyl CoA**. This is catalyzed by **propionyl CoA carboxylase**, which is a **ligase** because ATP is simultaneously converted to AMP and PP_i to power this reaction.
2. D-methylmalonyl CoA is converted to **L-methylmalonyl CoA** by a **racemase** enzyme, a type of isomerase.
3. L-methylmalonyl CoA is converted to **succinyl-CoA** by the enzyme **methylmalonyl-CoA mutase**. This enzyme requires **vitamin B₁₂** as its coenzyme. Once succinyl-CoA is formed, it can then enter into the Krebs cycle.

The Krebs cycle deserves special mention: not only does the Krebs cycle function in the oxidative catabolism of amino acids, fatty acids, and carbohydrates, but it also serves as a primary starting point in many biosynthetic reactions for which it is able to provide precursors. If these precursors were removed from the Krebs cycle for various other metabolic pathways, then the rate at which the Krebs cycle operated would begin to decline. One of the most important *anaplerotic* reactions is the synthesis of OAA from pyruvate and CO₂. The reaction is catalyzed by the enzyme **pyruvate carboxylase**. If a sample was *exclusively* catabolizing fatty acids to obtain a high level of ATP, and if there was no carbohydrate catabolism taking place, then the anaplerotic reaction of pyruvate to OAA would not take place. The result is that the Krebs cycle would run down. In other words, fatty acid degradation needs the ‘flickering flame’ of carbohydrate degradation to keep it going (i.e., a little conversion of pyruvate to OAA is needed to ensure that fatty acid degradation will continue).

Urea Cycle: If we hydrolyze proteins, we will be able to obtain a variety of amino acids, which can be used for protein synthesis. However, for most amino acids, the amino group can also be removed to yield the α -keto acids, which can then be used in citric acid cycle intermediates to eventually give ATP, CO₂ and H₂O.

Phase I essentially involves two reactions that mainly occur in the liver. The first is amino transfer from an amino acid to **α -ketoglutarate**, which is the acceptor molecule of the amino group from most amino acids. The enzyme that catalyzes this reaction is an **aminotransferase**. There is a specific aminotransferase for each amino acid. The mechanism of these reactions involves the coenzyme **pyridoxal phosphate (PLP)**, which is derived from **vitamin B₆**. The important takeaway from this reaction is that the amino acid loses its amine group.

When **α -ketoglutarate** accepts the amino group, it actually becomes the amino acid **glutamate**. This glutamate then gets oxidatively deaminated by **glutamate dehydrogenase**, and either NAD⁺ or NADP⁺ can be utilized. The nitrogen is released in the form of NH₄⁺. In terrestrial vertebrates, the excess ammonium produced is converted into **urea** and then excreted. In most aquatic organisms, NH₄⁺ itself is excreted, while in birds and terrestrial reptiles, the NH₄⁺ is converted to **uric acid** before it is excreted. This concludes Phase I.

Phase II involves degrading the α -keto acids to form metabolic intermediates which can be funneled into the Krebs cycle. Some of these amino acids will provide the carbon skeleton framework for the net synthesis of carbohydrate via **gluconeogenesis**. These amino acids are termed **glucogenic** and will be degraded to pyruvate, α -ketoglutarate, succinyl-CoA, fumarate, and oxaloacetate. Other amino acids will be degraded to acetyl-CoA and acetoacetyl-CoA. These amino acids are termed **ketogenic** because they will eventually give rise to **ketone bodies**.

As we have seen, amino groups will flow from the various amino acids to glutamate, and then to ammonia by the glutamate dehydrogenase reaction. The free ammonia that we produce in the glutamate dehydrogenase reaction combines with CO₂ to form **carbamoyl phosphate**, catalyzed by **carbamoyl phosphate synthetase**. This reaction requires 2 ATPs and occurs in the mitochondrial matrix. Carbamoyl phosphate then reacts with a molecule of **ornithine** (an amino acid that doesn't appear in proteins). **Carbamoyl transcarbamoylase** transfer the carbamoyl group from carbamoyl phosphate to ornithine to produce **citrulline**, which is another example of an amino acid that does not occur in proteins. Citrulline can leave the mitochondrial matrix and enter into the cytoplasm.

If we take citrulline and aminate it using **aspartate** via the enzyme **argino-succinate synthetase**, we will form **arginosuccinate**, which can be cleaved by the enzyme **arginosuccinase** to give **arginine** and **fumarate**. If you hydrolyze arginine with water, you will get ornithine and urea—urea can be excreted. This **urea cycle** is shown in Fig. 5.17.

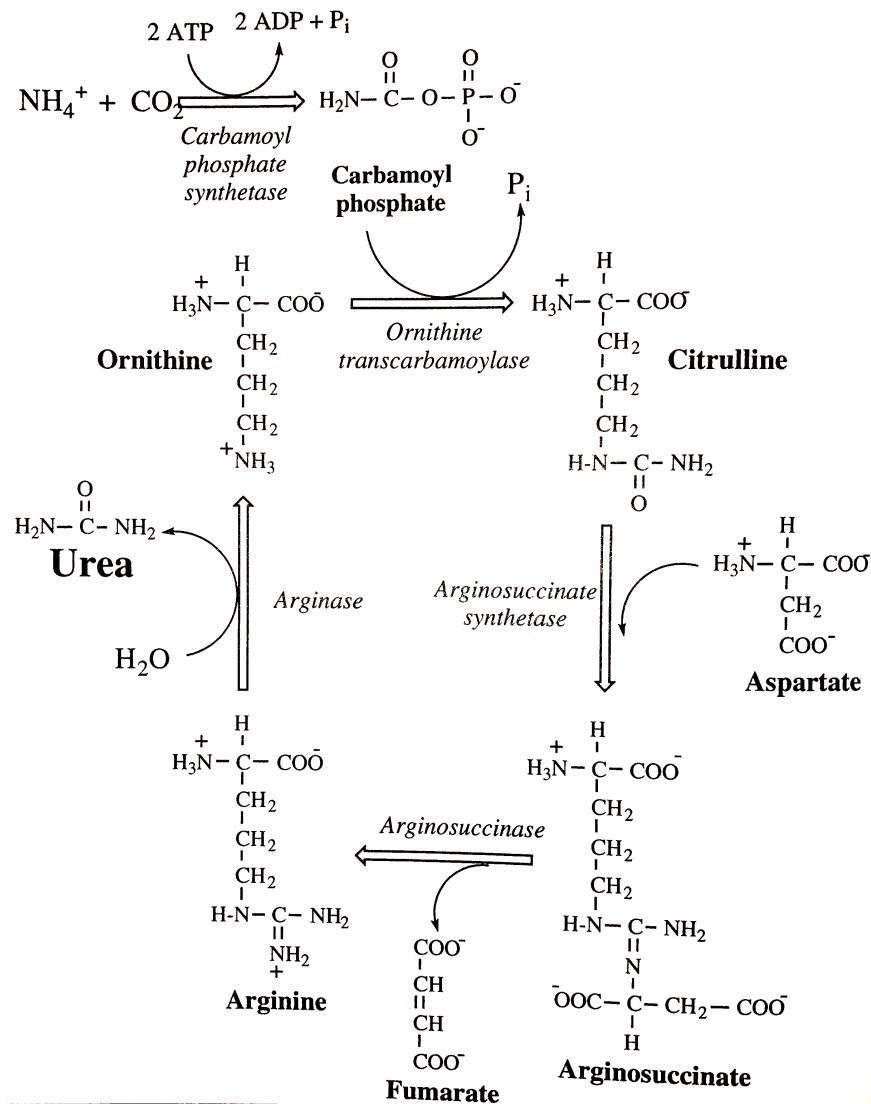


Figure 96: Overview of the urea cycle.

5.18 Genetic Information—Classical Genetics

Mendel's First Law of Heredity (also called the **Law of Segregation**) is that alternative alleles segregate from each other in heterozygous individuals and retain their identity. Mendel's Second Law of Heredity (also called the **Law of Independent Assortment**) is that the hereditary factors (i.e. genes) for different things assort independently of one another. Note that independent assortment of the genes will occur if they are located on different chromosomes or are far apart on the same chromosome.

5.19 Genetic Loci & Alleles

Tryptophan is considered an **essential amino acid** because it is an amino acid that an organism cannot synthesize itself, and therefore must be obtained from the diet. An *auxotroph* is a mutant that will grow only when its medium is supplemented with a particular compound which is not required by the normal wild type organism. The wild type organism is referred to as a *prototroph*. By definition, an auxotroph will not grow on a minimal medium, while a prototroph will.

Pleiotropy is when an individual allele has more than one effect on the phenotype (i.e. there's a mouse gene that determines both the fur coat color and the viability of the mouse). **Epistasis** is when multiple genes all contribute to a particular phenotype and are able to interact with one another. This occurs between different pairs of genes, *not* between two members of an allelic pair. For example, in order for tryptophan to be synthesized, the dominant allele of all the genes involved in this biosynthetic pathway must be present (in the absence of tryptophan), so the genes involved in the biosynthetic pathway of tryptophan are said to act in an epistatic fashion.

The short arm of a chromosome is called the *p* arm, while the long arm of a chromosome is called the *q* arm.

5.20 Pedigrees & DNA

In pedigree charts, males are squares and females are circles. If a square or a circle is filled in with a dark color, then that individual is affected with a given defect. If there is only one copy of any chromosome (i.e. **monosomy**), the individual will not survive development. The most common form of **trisomy** is for chromosome 21, which leads to Down's syndrome. It is an example of **aneuploidy** (the condition in which nuclei have an unbalanced set of chromosomes—that is, they do not contain an exact multiple of the haploid number of chromosomes).

Double stranded DNA has a lower absorbance (by about 40 to 50%) than single stranded DNA. Also, GC-rich DNA has a lower absorbance than AT-rich DNA at a given temperature. DNA replication is **semiconservative**. Adenine and guanine are purines (two rings in their structures), while thymine and cytosine are pyrimidines (one ring in their structures). Double stranded DNA is a right-handed helix.

5.21 DNA Synthesis

All of the dNTPs have a deoxyribose sugar backbone (β -D-2'-Deoxyribose) with an N-glycosidic linkage from the C-1 carbon to the purine or pyrimidine. There's also a phosphomonoester linkage from the phosphate group to C-4 carbon of the sugar backbone, and additional phosphate groups can chain together via acid anhydride linkages. All dNTPs have 3 phosphate groups chained together. **DNA polymerase I** adds about 20 dNTP residues to the 3'-hydroxyl function of a pre-existing DNA strand. The dNTPS are added to the newly synthesized DNA chain in the $5' \rightarrow 3'$ direction, which means the DNA template is read in the $3' \rightarrow 5'$ direction.

There are a couple different types of DNA structures that are helices: A-DNA, B-DNA, and Z-DNA. These aren't really important, just know that the canonical DNA form that we're all familiar with is the **B-DNA** with the major and minor grooves. In A-DNA, β -D-2-Deoxyribofuranose ring can pucker, and the minor groove essentially vanishes due to the puckering of the furanose ring. The type of helix that is found in A-DNA is also found in regions of double stranded RNA (e.g. involving hairpins) and in RNA-DNA hybrids. Z-DNA is actually a left-handed helix, and involves the phosphates in the backbone zigzagging due to a repeating dinucleotide unit, as opposed to a mononucleotide unit. In Z-DNA, when the pyrimidine bases and the ribose units are far apart, it is in an **anti** conformation, while it is in a **syn** conformation when the purine bases and the ribose units are close together.

Positive supercoiling is twisting a DNA molecule around its own axis in the right-handed direction, while negative supercoiling is twisting a DNA molecule around its own axis in the left-handed direction (these conventions are for right-handed DNA). The number of times that one DNA strand can be wound around another DNA strand is referred to as its **linking number (L)**. **Topoisomers** are DNA molecules that differ only in their linking number. The degree of the linking number in DNA can be altered by enzymes called **topoisomerases**. **Type I** topoisomerases reversibly cleave *one strand of DNA* and relaxes negatively supercoiled DNA, while **Type II** topoisomerases (e.g. DNA gyrase) reversibly cleaves *both strands of DNA* and adds negative supercoils. Negatively supercoiled DNA prepares the DnA duplex for processes like replication, recombination, and transcription where the separation fo the two helical strands is required.

The DNA replication process is very systematic. The synthesis of new daughter DNA is coupled to the unwinding of parental DNA at sites called **replication forks**, which occur at sites referred to as **origins** (ori for short). In order for the DNA double helix to unwind, **DNA gyrase** adds negative supercoils ahead of the advancing replication fork, as when the replication fork is initially unwound, positive supercoils are introduced making DNA separation rather difficult without the DNA gyrase activity. **Helicase** then binds to the ori site and catalyzes the ATP-driven unwinding of the duplex DNA.

Single-stranded binding proteins then stabilize the unwound portion of the parental DNA.

From here, both strands can serve as template strands, but DNA Polymerase I can only synthesize DNA in the $5' \rightarrow 3'$ direction. We therefore have a leading strand with continuous replication, and also a lagging strand that uses **Okazaki fragments** to discontinuously synthesize the complementary strand (the primers used are RNA primers). DNA Polymerase III holoenzyme is used to synthesize the lagging strand (as opposed to DNA Polymerase I) because it is quicker and can add more dNTP's to the growing DNA strand. DNA polymerase I also has $5' \rightarrow 3'$ exonuclease activity that allows it to remove the short segments of RNA primer, and then add deoxyribonucleotides to the free 3'-hydroxyl function of the chain undergoing elongation. **DNA ligase** then joins together the Okazaki fragments. It is important to remember that replication of DNA can be bidirectional.

5.22 Expression of Genetic Information

DNA polymerase uses dNTP's (releasing PP_i) while RNA polymerase uses NTP's (also releasing PP_i). DNA polymerase also requires a primer with a free 3'-OH group to initiate chain elongation. RNA polymerase does not require a primer.

In prokaryotes such as *E. coli*, there is only one RNA polymerase that has to make all the types of RNA in the cell, including transfer RNA (tRNA), ribosomal RNA (rRNA), and messenger (mRNA). What tells an RNA polymerase where to start transcription? It turns out that there are very specific starting sites for transcription which are different in both prokaryotic and eukaryotic organisms.

The signals that tell the RNA polymerase where to begin transcription are upstream from the +1 site. In a prokaryotic organism, these areas are usually centered around the -10 region (TATAAT) and the -35 region (TTGACA). The -10 region is referred to as the **Pribnow box** while the -35 region is referred to as the **-35 region**. As we will see, there is some flexibility in these numbers from gene to gene or organism to organism. Therefore, we can speak of these regions as being **consensus sequences**. As the RNA polymerase moves along the DNA duplex it is looking for the proper signals that will tell it where to promote transcription. Hence, these regions are called **promoter sites**.

In the case of eukaryotic organisms, the promoter areas are centered around a -25 region (TATA box) and a -75 region (CAAT box) upstream from the +1 site. The -25 region is referred to as the **Hogness box**. The CAAT box may or may not be present, and may even be moved further upstream. There are also **enhancer sequences** which can help stimulate transcription. These enhancer sequences are usually quite far away from the promoter region (e.g. several kb) and can be either upstream or downstream from the promoter region. The process of whether or not a gene gets transcribed no only depends on the correct promoter region, but also on the interaction with specific enhancers.

In prokaryotes, the RNA polymerase will proceed down the DNA template until it runs into the terminator sequences, which is a base-paired hairpin sequence. Following this hairpin structure is a sequence of four or more U residues (referred to as the **poly-U tail**). Once this hairpin pairing occurs in the RNA molecule, the RNA polymerase pauses, and the RNA chain will dissociate from RNA polymerase and the DNA duplex. This process is referred to as **rho-independent termination**. The second way in which termination can be accomplished involves the ρ protein and is referred to as **rho-dependent termination**. The ρ complex can specifically bind newly synthesized ssRNA and pull itself towards the replication bubble where it will dislodge RNA polymerase from the DNA template.

In the case of eukaryotic cells, there are three separate types of RNA polymerases. **RNA polymerase I** will transcribe rRNA, **RNA polymerase II** will transcribe mRNA, and **RNA polymerase III** will transcribe tRNA.

The prokaryotic (specifically *E. coli*) **RNA polymerase** is composed of 4 subunits: α_2 , β , β' , and σ . Together, these four subunits are called the **RNA polymerase holoenzyme**. The σ subunit is responsible for locating the promoter site that initiates transcription. Once RNA synthesis begins, the σ subunit will dissociate from the RNA polymerase holoenzyme. RNA polymerase, which now has three subunits ($\alpha_2\beta\beta'$) is called the **core enzyme**. The core enzyme can still initiate transcription even though it has lost its σ subunit, but it does so poorly. Because the holoenzyme has lost one of its subunits (the σ subunit), it is called an **apoenzyme**—an ‘incomplete’ enzyme. As we can see, σ factors are quite important in bacterial transcription, and is an example of a regulatory subunit.

Recall that during DNA replication, there were two processes that ensured the accuracy of the final product: (1) editing by DNA polymerase enzymes and (2) the repair enzyme processes. In contrast, the RNA polymerases do *not* have a polymerase editing function and they don't have an enzyme repair system. As a result, the probability of an error occurring is about one in every 10^4 to 10^5 transcribed bases, roughly 5- to 6-fold higher than the rate in DNA replication. These higher error rates are fine because we aren't making proteins that are 10^4 amino acids long.

Following transcription, a variety of modifications to RNA transcripts can occur. There can be (1) trimming and splicing, (2) addition of bases to the primary transcript, and (3) modification of bases, a feature that is particularly evident with tRNA molecules. Processes (2) and (3) are quite common in prokaryotic cells. In eukaryotic cells, there are other addition

modifications:

1. **5' Capping of the Nascent RNA:** To prevent degradation of the RNA transcript by phosphatases and nucleases at the 5' end, this end is modified by ‘capping’ it with a GTP-derived cap. A methyl donor **S-adenosylmethionine (SAM)** also adds a methyl group to the guanine nucleotide to complete the capping process.
2. **3' Polyadenylation of the Nascent RNA:** As eukaryotic RNA polymerase nears the end of its transcription process, about 250 adenine residues are added to the free 3' end of the RNA polymer to form a **Polyadenylated tail**, which helps to protect the 3' end of the RNA polymer from nucleases and phosphatases.
3. **Splicing:** Introns are spliced out of the primary RNA in a process called **splicing**, and the exons are joined together. There are self-splicing RNA molecules, which was the first clear example in which proteins and coenzymes are *not* the only catalysts—RNA could be involved in its own chemistry without the help of any enzymes.

Carboxypeptidase is a digestive enzyme that hydrolyzes the *C-terminal* peptide bond in polypeptide chains, particularly if that residue has a bulky aliphatic side chain or an aromatic ring. In general, peptide bond formation between two amino acids is thermodynamically unfavorable—this unfavorability is overcome by activation of the carboxyl group of incoming amino acids. There is a serial, step-wise process on how this occurs:

1. ATP reacts with an amino acid to form an activated, mixed anhydride linkage in the form of an **aminoacyl-AMP**.
2. The aminoacyl group of aminoacyl-AMP is transferred to a tRNA molecule to form **aminoacyl-tRNA**, which is the *activated* intermediate in protein synthesis. The enzyme that catalyzes this reaction is an activating enzyme (i.e. a specific aminoacyl-tRNA synthetase).

There has to be great specificity regarding how the amino acids are placed on the different tRNA molecules in the cell. This is achieved by the 20+ activating enzymes present—one for each amino acid. Each activating enzyme is quite selective for its particular amino acid.

All tRNA molecules are capable of forming a cloverleaf structure that is about 80 ribonucleotides long. An amino acid would be attached to the terminal adenosine residue at the 3' end of the tRNA. The 5' end of the tRNA has a free phosphate group. tRNA has certain secondary structures, and one of the more important ones is the **anticodon loop**, which is complementary to the codon of the mRNA that calls for a particular amino acid. Note that the codon-anticodon relationship is anti-parallel. Sometimes, a tRNA base can pair with two different codons—the basis for this phenomenon has to do with **wobble** in the base pairing. To introduce wobble, let's start off with an example.

There are two codons—5'-UUU-3' and 5'-UUC-3'—that are both recognized by the same tRNA that brings in a phenylalanine residue. The activated Phe-tRNA^{Phe} has an anticodon loop of 3'-AAG-5', which is the complement of only one of the codons. So we have a standard base pairing and a non-standard base pairing. In the case of the non-standard base pairing between the U and the G, the pyrimidine base is displaced with respect to the purine base. If this base pairing were to occur in a DNA double helix, it would distort the backbone, but in tRNA this doesn't matter too much because we are not making that large of a double helical structure. We simply have three bases of mRNA pairing with three bases of tRNA. In other words, there is a great deal more flexibility in this type of interaction in tRNA compared to DNA.

As this example demonstrates, **wobble** is simply the steric freedom in the pairing of the third base of the codon. The wobble rules apply at the first base of the anticodon and the third base of the codon. These rules are:

1. U can pair with A (standard) or G (non-standard).
2. G can pair with C (standard) or U (non-standard).
3. Inosine (I) can pair with either A, U, or C. **Inosine** is commonly found in tRNAs and allows for a rather large degree of flexibility in the third base pair.

A prokaryotic ribosome (**70S**) can be dissociated into a small subunit (**30S**) and a large subunit (**50S**). Note that there is a non-linear relationship size and the Svedberg unit sedimentation coefficient value. The dissociation of the 70S ribosome is controlled by the concentration of Mg²⁺—low Mg²⁺ causes dissociation of the ribosome.

In addition to the aminacyl-tRNAs, mRNA, and ribosomes for protein synthesis, we also need additional proteins called **initiation factors** that are not part of the ribosome. In general, there are initiation, elongation, and termination factors. GTP would also be required as an energy source.

In prokaryotes, about 10 nucleotides upstream from the initiating codon (5'-AUG-3') is a purine (AG)-rich sequence of

bases called the **Shine-Dalgarno sequence**. This sequence is responsible for binding the mRNA to the **16S subunit**—an architectural constituent of the 30S ribosomal subunit. Binding of the Shine-Dalgarno sequence to the 3' end of the 16S ribosomal subunit forms the **initiation complex**.

After the 30S subunit has formed the initiation complex with the appropriate initiation factors, GTP will bind to a particular initiation factor called **IF2**. This allows the mRNA and the **fMet-tRNAsf** initiator signal to join the 30S subunit as well. fMet is a modified methionine residue, created when a formyl group attached to THF reacts with Met-tRNAsf to form **formyl-methionine-tRNAsf**, abbreviated as fMet-tRNAsf, and THF without the formyl group. The subsequent release of these initiation factors allows the 50S complex to attach to the 30S complex, thus forming the 70S initiation complex. The 70S initiation complex is now primed for protein synthesis.

On a side note, it turns out that there are two tRNAs for methionine. One form of tRNA, called **tRNAsf**, has the initiation role we just mentioned, while the other rRNA, called **tRNAm**, will simply bind to AUG codons elsewhere in the mRNA.

The ribosome has two sites for aminoacyl-tRNAs, the A-site and the P-site (this is shown in the 70S initiation complex in **Fig. 5.22**). Immediately after initiation, the P-site is occupied by fMet-tRNAsf. Let's assume that the next amino acid that we want to add is arginine. Then the activated Arg-tRNA^{Arg} would enter the A-site, and the enzyme **peptidyl transferase** catalyzes the formation of a peptide bond. The fMet moiety of the P-site will be transferred to the amino group of the Arg residue of the A site, and the formation of the dipeptidyl-tRNA at the A-site is a favorable reaction.

The empty tRNAsf will leave the P-site and the mRNA will move a distance of three nucleotides, thus bringing the

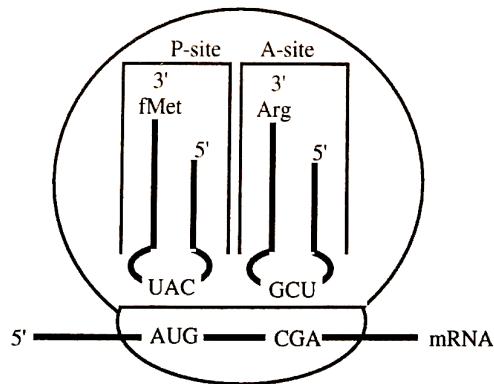


Figure 97: 70S initiation complex. Note the order of the P-site and the A-site.

dipeptidyl-tRNA (5'-fMet-Arg-3' in this case) to the P-site. This process is called **translocation**. We are now left with the empty A-site and are ready for the next incoming amino acid during the next round of elongation. Adding all the amino acids except for the first fMet one takes 4 ATPs of energy—2 for activation, 1 for binding at the A-site, and 1 for translocation to the P-site. The first fMet one only takes 3 ATPs of energy because it doesn't need to translocate.

Finally, let's consider the termination process. Suppose we are close to the end of the 3' end of the mRNA. A stop codon termination signal is *not* recognized by a tRNA, but instead are recognized by a **protein release factor** that will bind to the signal when it comes into view. This release factor has an effect on the peptidyl transferase enzyme—, namely, it will now use water instead of an amino group to attach the amino acid at the P-site, thus releasing the polypeptide chain.

Puromycin is an antibiotic that binds at the A-site and acts as an analog of an aminoacyl-tRNA, thus preventing other aminoacyl-tRNAs from entering the A-site. The α -amino group of puromycin forms a peptide bond with the carboxyl group of the previous amino acid at the P-site. Once this happens, *peptidyl puromycin* leaves the 70S ribosome, thus causing premature termination of protein synthesis.

The Lactose Operon: An **operon** is simply a transcription unit which is involved in the expression of multiple genes. Another way to word this would be that an operon is an arrangement of genes on the DNA which are co-transcribed.

The lactose operon includes a **promoter (P)**, an **operator (O)**, and **structural genes (z, y, and a)**. Upstream from the lactose promoter is the **regulatory gene (i)** and its own **promoter (Pi)**. Structural gene **z** codes for β -galactosidase whose major function is to hydrolyze lactose to glucose and galactose. Structural gene **y** codes for **galactoside permease**, a carrier molecule that transports lactose into the cell. Structural gene **a** codes for **thiogalactoside transacetylase** with unknown function.

The **regulatory gene (i)** codes for a repressor protein that binds to the **operator** and prevents the transcription of

genes z, y, and a when lactose is not present. This system allows the cell to not produce unnecessary proteins encoded by genes z, y, and a if there's no lactose around. However, once lactose is present in the medium, then a small amount of it will be converted to **allolactose** by a few molecules of β -galactosidase that are present in the cell—allolactose binds to the repressor and decrease its affinity for the operator site. Once this *inducer* binds to the repressor, the repressor-inducer complex dissociates from the operator and RNA polymerase can begin transcription of the lactose operon genes.

As another layer of complexity, the transcription of *E. coli* lactose operon also depends on the relative concentrations of glucose in the cell's medium. The idea is that when glucose is present in the medium, there is a preferential reduction in the synthesis of the enzymes needed to utilize lactose, even though lactose may be present in high concentrations—this is called **catabolite repression**. A considerable distance downstream from the structural genes was the *crp* gene that codes for the **catabolite activating protein (CAP)**, which is also called the **cAMP receptor protein (CRP)**. CAP is a DNA binding protein that has the ability to mediate catabolite repression of many inducible operons, such as the lactose operon, *only if* cAMP is complexed with it. cAMP is synthesized from ATP and increases in concentration when glucose levels are low. Thus, when cAMP binds to CAP, it forms the active cAMP-CAP complex which can then bind to the DNA at the CAP site, located just upstream of the lactose promoter (see **Fig. 5.22**). The cAMP-CAP complex acts as an activator and binds to the CAP-cAMP site. Transcription of the lactose operon is enhanced by somehow converting a weak lactose promoter into a stronger lactose promoter.

The Tryptophan Operon: The tryptophan (*trp*) operon has five structural genes: *trpA-trpE*. These are essential

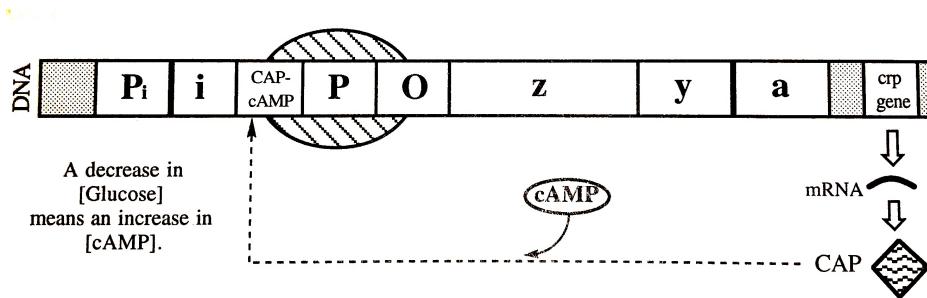


Figure 98: The lac operon in prokaryotes.

for the prokaryote to be able to synthesize its own tryptophan amino acid, and mutants with structural gene mutations have to have the amino acid supplemented in their diet. There is also the *trpR* gene, which codes for the **tryptophan repressor**. In order to regulate the synthesis of tryptophan in a negative feedback loop, the repressor must bind the end product of the metabolic pathway of interest, which is tryptophan itself. Therefore, tryptophan acts as a **co-repressor** for its own biosynthesis.

Upstream from the start of the structural genes and downstream from the operator is a sequence called the **leader sequence**. This sequence is coded for by the *trpL* gene. Towards the 5' end of the leader sequence is a region that codes for 14 amino acids called the **leader peptide**. The 10th and 11th amino acids of this leader peptide are the **tryptophan residues**. These two amino acids are also part of a nucleotide sequence that is *GC-rich*. There are four GC rich regions in the leader sequence, and are sometimes called region 1, region 2, region 3, and region 4, from upstream to downstream order. Because these GC regions are complementary to each other, we find two alternative secondary structures resembling hairpins can be formed—one between regions 1 and 2 and another between regions 3 and 4. This represents a *p-independent* termination site.

Given this information, let's examine transcription of the leader region in a little more detail. Suppose that the cell has a sufficient level of tryptophan. As RNA polymerase transcribes region 1 and 2, these two regions begin to form a hairpin structure that causes the polymerase to pause. By this time, the Sine-Dalgarno sequence towards the 5' end of the mRNA has been made, and a ribosome binds and begins translation. As the ribosome translates the leader peptide region, it disrupts the hairpin created between regions 1 and 2.

RNA polymerase then begins to transcribe region 3. Meanwhile, the ribosome reaches the tryptophan codons from the structural genes and inserts the required tryptophan amino acids into the growing polypeptide chain. As the ribosome moves downstream a few more codons, it encounters a termination codon and stops translation. The ribosome dissociates and the hairpin between regions 1 and 2 reform. As RNA polymerase finishes transcription of region 4, a hairpin structure forms between regions 3 and 4. This hairpin is a termination signal because of the *ribo-U's* that immediately follow region 4. Once this termination signal is formed, the transcription complex dissociates from the duplex DNA and structural genes *are not*

transcribed. Thus, tryptophan will *not* be synthesized.

Now, suppose the levels of tryptophan are really low. When the ribosome reaches the tandem tryptophan codons, it *pauses* because the concentrations of tryptophan in the cell are low, and so there's not enough charged tryptophanyl-tRNA^{Trp} molecules to bring tryptophan to the site of protein synthesis. This results in the ribosome *sitting on* region 1 and covering it up. Meanwhile, RNA polymerase has transcribed region 2 and region 3. Since region 1 is not available for hairpin formation with region 2, region 2 instead elects to form a hairpin structure with region 3. By this time, RNA polymerase has transcribed region 4 and the string of ribo-U's.

Region 3 will not base pair with region 4 because the hairpin structure formed between regions 2 and 3 is more stable, so the termination hairpin is *not* formed. The transcription complex is not disturbed and RNA polymerase transcribes the structural genes of the *trp* operon. This type of regulation of the synthesis of mRNA is referred to as **attenuation**, and the control element which is responsible for this phenomenon is called an **attenuator**. All of this information is shown in Fig. 5.22.

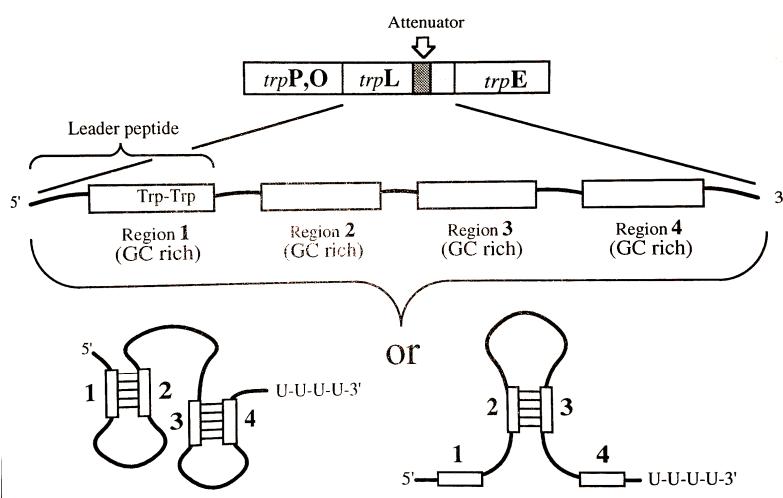


Figure 99: The partial *trp* operon in prokaryotes. We show the leader sequence and resulting hairpins formed when the concentration of tryptophan in the cell is both high (left) and low (right).

5.23 Mutations

Substitution mutations can be divided into two types: transition mutations and transversion mutations. **Transition mutations** occur when one purine is replaced by another purine or one pyrimidine is replaced by another pyrimidine. **Transversion mutations** occur when one purine is replaced by a pyrimidine or one pyrimidine is replaced by a purine.

Transition mutations can occur between certain hydrogen atoms on each of the four bases can change their positions, thus giving rise to **tautomeric structures**. For example, consider the pyrimidine base **thymine** (in the **lactam** form). Under normal conditions, thymine will bond with adenine. However, if the hydrogen atom on the nitrogen at position 3 in the thymine ring moves to the oxygen atom at position 4, then we have a *rare* form of thymine (the **lactim** form) that is capable of hydrogen bonding with **guanine**. These structures are shown in Fig. 5.23. The fraction of thymine in the rare lactim form is about 10^{-4} . This tautomerization can result in transition mutations, as shown in Fig. 5.23.

To preserve the fidelity of the genetic message, both **DNA polymerase I** and **DNA polymerase III** possess $3' \rightarrow 5'$ exonuclease activity. These polymerases are continually checking the segment of new DNA that is newly synthesized—if it detects an error, then that error will be removed by the $3' \rightarrow 5'$ exonuclease activity of the enzyme. After the correction is made, the polymerase will continue with the replication of the DNA template.

Recall that the bases cytosine, adenine, and guanine have *amino groups* on them that can be hydrolyzed. For example, cytosine can be deaminated to form **uracil**, which is analog of thymine that pairs with adenine. (Nitrous acid HNO_2 is actually a mutagen that can cause the deamination of cytosine to uracil, resulting in a transition mutation.) If a cytosine were to be deaminated to form uracil in the template strand of DNA, then the polymerase would put in an adenine at the corresponding position instead of a guanine. This is another way in which transition mutations can arise.

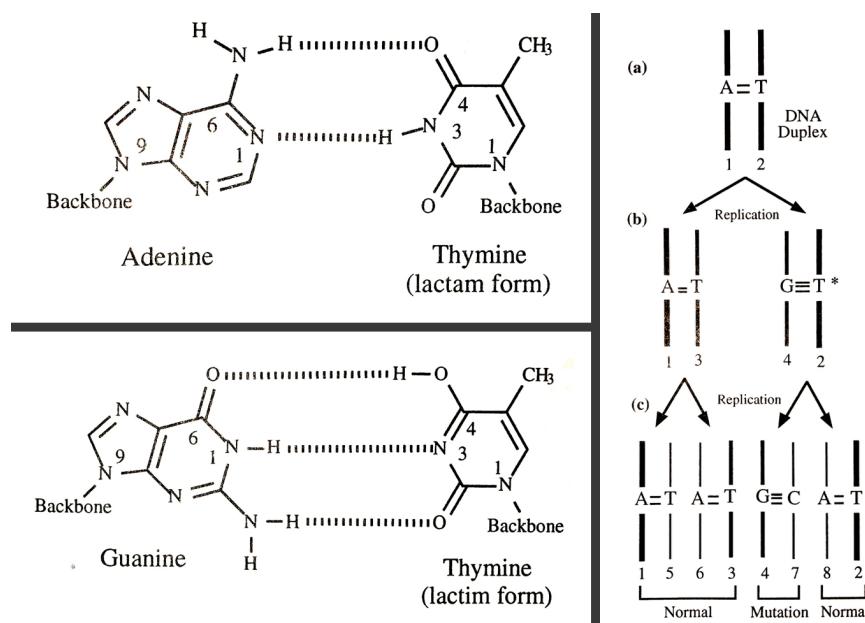


Figure 100: *Left, upper* Normal pairing between A and T, with thymine in the lactam form. *Left, lower* Tautomeric conversion of T into the lactim form, resulting in abnormal pairing between G and T. *Right* Demonstrating how tautomerization of thymine can give rise to transition mutations.

Luckily, the cell has a repair system that recognizes uracils and removes them. The enzyme **uracil-DNA glycosidase** hydrolyzes the bond between the deoxyribose ring and the uracil base, thus removing the uracil base. This site on the DNA duplex is called an **AP site** (either apurinic or apyrimidinic), because it is without either a purine base or a pyrimidine base. This base defect is recognized by an enzyme called **AP endonuclease** which cleaves the bond on the 3' side of the phosphodiester bond of the nucleotide with the missing base. **DNA polymerase I** cleaves the phosphodiester bond at the 3' end, and then will read the complementary strand to insert the correct base at the AP site on the damaged DNA. **DNA ligase** will then seal the inserted base into the damaged strand. **Mismatch Repair:** The most common types of mutations are caused by tautomerism and slippage during replication. (Slippage is when a base ‘slips out’ of register from DNA polymerase III, resulting in insertion and deletion mutations.) However, even with our proofreading system, there could be errors where a base is paired with the wrong complementary base (i.e. a ‘mismatch’) that is missed by the DNA polymerase—because of the resulting imperfect base pairing interaction, there is a slight bulge in the DNA duplex at that point. The **mismatch repair enzyme** slides along the DNA duplex ‘feeling’ for bulges like these. However, even when the enzyme finds this bulge, it doesn’t know which base was incorporated into the replicating DNA in error, so it needs a way to determine which strand has the ‘legit’ sequence and which is the error sequence.

The way that it accomplishes this is by searching for the palindromic sequence GATC in the DNA, referred to as a **GATC box**. This box can tell the enzyme which is the newly synthesized strand—the parent strand has the *adenine* base methylated in each of these GATC boxes. The methyl groups are attached to adenine by **SAM (S-Adenosylmethionine)**. This process of methylation is a signal that identifies the old strand from the new strand, as the newly synthesized DNA strand is not methylated until a few seconds/minutes after the replication process. Thus, once the incorrect base is located, the mismatch repair enzyme binds to the unmodified GATC box (with the non-methylated adenine) *and* to the incorrectly inserted base, turning/contorting the DNA as necessary to do this. The mismatch repair enzyme then removes the mispaired base and the intervening sequence of DNA up to the point of the unmethylated GATC box. **DNA polymerase I** will then resynthesize the DNA that has been removed by the mismatch repair enzyme, and DNA ligase then patches it up.

If two pyrimidine residues such as thymine are adjacent to one another, then the energy stored in UV radiation can link them together—the resulting **thymine dimers** inhibit DNA replication. Luckily, these pyrimidine dimers can be corrected by a repair mechanism. A specific **endonuclease** enzyme called UvrABC will cut on both side of the dimer and remove the affected segment of DNA about 12 nucleotides in length; the resulting gap is filled in by DNA polymerase I and patched up with DNA ligase.

Of note is the tricyclic benzenoid hydrocarbon called **anthracene**, which has a basic planar structure due to its ring

system, similar to those on the DNA bases. Thus, anthracene can very easily *intercalate* between adjacent base pairs. Since anthracene is about the size of a base pair, they are quite capable of causing frame shift mutations as a result during the replication process.

5.24 Genetic Engineering & Cloning

Bacteria cells have a variety of restriction enzymes which can degrade the foreign DNA of an attacking virus. The bacteria's restriction enzymes do not degrade its own DNA because its DNA has been methylated at specific locations by modification enzymes. When the viral DNA is injected into the bacterial cell, its DNA has not been methylated yet and so the bacteria's restriction enzymes recognize that DNA as being foreign. Restriction enzyme sites are almost always palindromic.

The **Sanger dideoxy technique** is used for sequencing DNA, and makes use of the fact that DNA polymerase I can be inhibited by a 2',3'-dideoxynucleoside triphosphate (ddNTP) analog of one of the four bases. The process is as follows:

1. Select the DNA strand to be sequenced. Attach a short radioactively labeled section of primer to the DNA strand being analyzed. The label of choice is ^{35}S instead of ^{32}P because ^{35}S has a longer half life. Also, ^{35}S has such a weak emission that it doesn't require shielding. The ^{35}S replaces the oxygen atom that is double bonded to the α -phosphate of dNTP.
2. Create four solutions, all of which have DNA polymerase I, the four dNTPs, and *one* of the ddNTP's in a controlled ratio.
3. DNA polymerase I will take the dNTPs and add them to the primer. The 2',3'-dideoxy analog can be incorporated into the growing DNA chain, but when the next dNTP is used in the reaction sequence it cannot be added to the 2',3'-dideoxy analog because the analog does not contain a 3'-OH function to form a phosphodiester bond. Termination of the DNA chain occurs specifically where the nucleotide analog was incorporated. This yields DNA fragments of different lengths containing the analog at the 3' end.
4. The labeled fragments are separated by size on a PAGE gel, and the pattern of fragments from the autoradiogram yields the DNA sequence that was synthesized using the radioactively labeled complementary DNA primer strand—not the original DNA strand you wished to analyze.

Fig. 5.24 shows a picture of a typical result from the Sanger sequencing method.

Imagine that we're trying to clone a fragment into a plasmid using restriction enzymes. You have to take into account

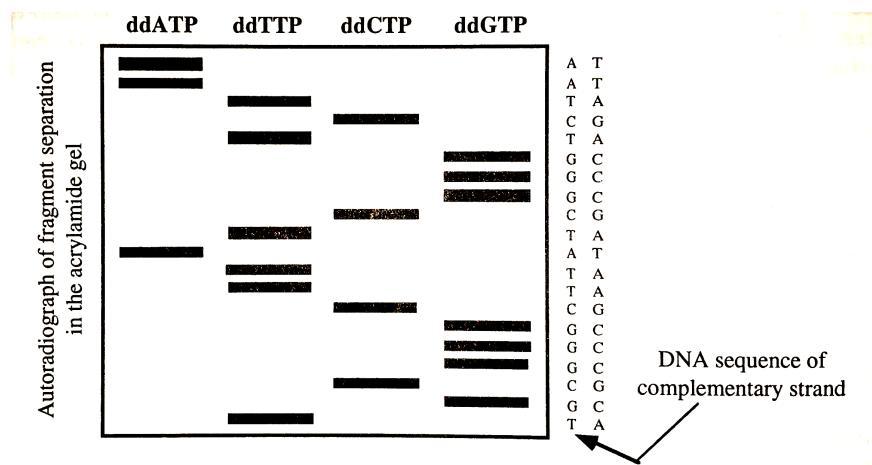


Figure 101: An example of the pattern of fragments from an autoradiogram for Sanger sequencing.

the copy number and size of the plasmid. Another problem is with reclosure of the DNA—you don't want the backbone to autoligate. Aside from adding excess fragment, another way to get around this reclosure problem is to treat the backbone with **phosphatase** to remove the 5' phosphate on the DNA duplex, which is absolutely required in order for DNA ligase to work. This prevents DNA ligase from closing the plasmid back up from a nick between nucleotide residues using ATP or NAD^+ . (In bacterial cells, the energy source is NAD^+ , while in animal cells, the energy source is ATP.) Thus, the plasmid

can only circularize with DNA ligase if the insert (with its 5' phosphate groups) is inserted into the backbone.

Another problem is that since restriction digestion sites are often palindromic, the insert can be inserted in either orientation, but only one of them is usually valid—the other orientation would produce some garbage mRNA. To confirm whether the insert was inserted correctly, you can do a confirmatory double restriction enzyme digest on the plasmid that produces asymmetric DNA fragment sizes depending on the possible orientations of the inserted DNA.

Cloning vectors include high-copy pUC plasmids (about 90 copies/chromosome) that you already know about and the **lambda (λ) phage**. This phage infects the bacterium *E. coli* and injects its linear DNA into the host. The viral DNA soon becomes circular and is quickly presented with two choices: (1) to become part of the host (the **lysogenic** pathway) or (2) to produce more progeny and destroy the host in the process (the **lytic** pathway). Certain portions of the phage's genome can be removed and foreign DNA inserted. Once the bacterium has been infected, the fragments of foreign DNA can be amplified many fold, thus giving rise to a **genomic library**. This library is then analyzed to find the gene of interest.

6 Psychology

6.1 The Biology Behind Behavior

Glial cells (glia) provide protection and support for neurons, maintain homeostasis, and form **myelin**, a fatty insulation wrapping axons to increase the speed and efficiency of signal conduction. Degeneration of myelin can result in *multiple sclerosis*. There are four main types of glia:

1. **Oligodendrocytes** wrap myelin extensions around the axons of neurons in the central nervous system (CNS).
2. **Schwann cells** wrap myelin extensions around the axons of neurons in the peripheral nervous system (PNS).
3. **Microglia** react to damage or disease as a macrophage would—they multiply, trigger inflammatory responses, and engulf cellular debris.
4. **Astrocytes** are the largest type of glia, and vary greatly in physiology and function. Some astrocytes cover blood vessels throughout the brain and are in contact with the cell bodies or neurons. These astrocytes help regulate the passage or prevention of certain chemicals between the blood and CNS neurons. Other types of astrocytes have been known to modulate neuronal activity, control the BBB, and to establish and sustain effective synapses.

A key difference between oligodendrocytes and Schwann cells is that oligodendrocytes often provide myelin to multiple axon segments on multiple neurons, but Schwann cells provide myelin to only one segment of a single neuron. Another important difference is that only Schwann cells can direct axonal regeneration after damage. This is why humans can recover from PNS nerve damage in the body, but damage in the brain or spinal cord frequently results in paralysis. The differences in oligodendrocytes and Schwann cells are shown in **Fig. 6.1**.

There are five major types of small-molecule neurotransmitters:

1. **Acetylcholine (ACh)** is the neurotransmitter present at every junction between motor neurons and skeletal muscles. ACh uptake by muscle cells causes those muscles to contract. ACh is also released in the CNS and has been demonstrated to play a role in learning and memory.
2. **Amino acid neurotransmitters** comprise the majority of fast-acting neurotransmitters found in the CNS. The four primary ones are *aspartate*, *glutamate*, *glycine*, and *γ -aminobutyric acid (GABA)*. Within the CNS, glutamate is the major excitatory neurotransmitter, and GABA (a modified derivative of glutamate, is the major inhibitory neurotransmitter.
3. **Monoamine neurotransmitters** are synthesized from a single amino acid and are slightly larger than the amino acid neurotransmitters. **Dopamine**, **epinephrine**, and **norepinephrine** are synthesized from the amino acid *tyrosine* and are all *catecholamines*. Neurons that release epinephrine are called *adrenergic* and neurons that release norepinephrine are called *noradrenergic*. The fourth monoamine neurotransmitter, **serotonin**, is synthesized from the amino acid *tryptophan* and is an *indolamine*.
4. **Unconventional neurotransmitters** include soluble-gas neurotransmitters (i.e. CO and NO) with high lipid solubility and thus fast response rate, and **endocannabinoids** that are similar to the primary psychoactive component found in marijuana (THC). Unlike most neurotransmitters, endocannabinoids are generally released from the dendrites and cell body of a neuron, not the axon terminal.

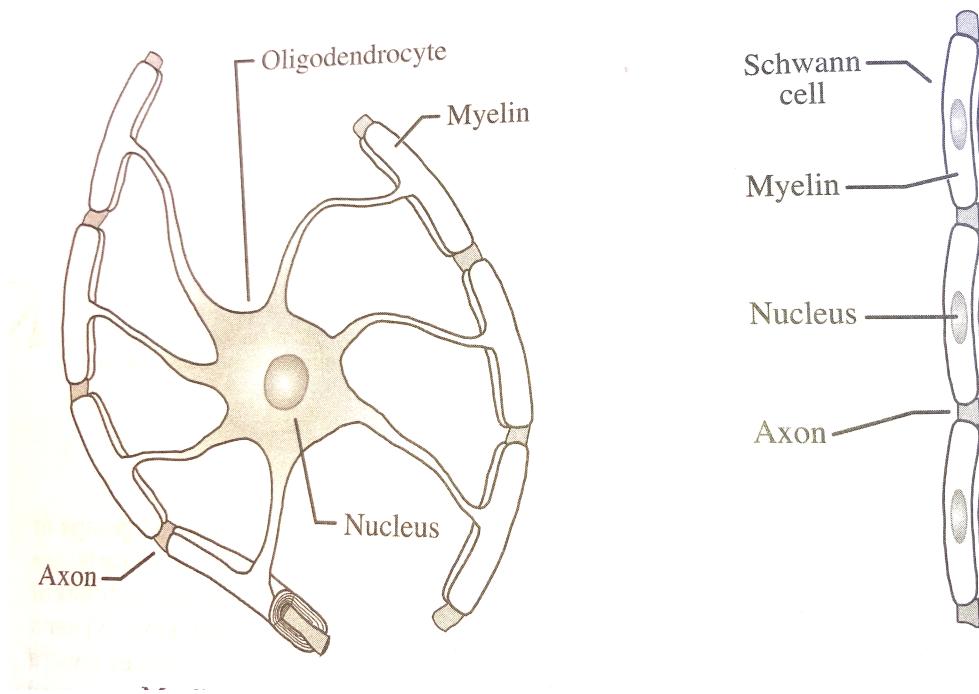


Figure 102: The difference in myelin wrapping between oligodendrocytes (left, CNS) and Schwann cells (right, PNS).

5. **Neuropeptides** are large-molecule neurotransmitters, and are separated into five categories: pituitary peptides, hypothalamic peptides, brain-gut peptides, opioid peptides, and miscellaneous peptides. They are used for hormonal communication. Neuropeptides are released from neuronal cells and signal to other neighboring neurons, meaning that communication is local. In contrast, peptide hormones are instead released from neuroendocrine cells and travel through the blood to various receptors in the body, meaning that communication is distant.
6. Another important class is **endorphins**, which are neuropeptide transmitters released in response to pain and intense exercise.

Neurotransmitters, drugs, and other chemicals act on synapses either to excite or inhibit neuron firing. As demonstrated in the similarity between endorphins and morphine, some molecules are structurally similar enough to certain neurotransmitters to allow them to bind to the same receptors and elicit a similar response—such molecules are called **agonists**. Other molecules that also bind to the same receptors but instead block the neurotransmitter's function are called **antagonists**. The functional divisions of the nervous system are shown in Fig. 6.1. The major anatomical structures of the brain are shown in Fig. 6.1.

The **myelencephalon**, also called the **medulla**, lies at the posterior (bottom) part of the brain and connects to the spinal cord. Of psychological interest is the **reticular formation**, a complex neural network of nuclei extending from the myelencephalon to the mesencephalon. The reticular formation plays an important role in regulating sleep, movement, attention, alertness, and various autonomous functions including heart rate, blood pressure, vomiting, coughing, and sneezing.

The **metencephalon**, also housing the reticular formation, contains two major structures: the pons and the cerebellum. The **pons** is a bulged section of the reticular formation that connects the medulla with the cerebral cortex, and also serves as a communication and coordination center between the body and hemispheres of the brain. The **cerebellum** is a large, convoluted structure that controls balance, coordination, equilibrium, and fine motor skills. Damage to the cerebellum does *not* result in paralysis, but rather disorders in fine movement and posture.

The **mesencephalon**, also housing the reticular formation, is divided into two sections: the tectum and the tegmentum. The **tectum** contains the **inferior colliculi**, regulators of auditory function, and the **superior colliculi**, regulators of visual-motor function. The **tegmentum** contains the **periaqueductal gray**, analgesic mediator of opiate drugs, and the **substantia nigra** and **red nucleus**, both important regulators of visual-motor function. Loss of the substantia nigra results in Parkinson's disease.

The **diencephalon** is divided into the thalamus and the hypothalamus. Acting as an important relay station to the cortex,

| Neurotransmitter | Function | Behavioral and Physical Malfunctions |
|---------------------|--|---|
| Acetylcholine (ACh) | Muscle stimulation, learning and memory | Motor control loss and dementia in Alzheimer's disease characterized by deterioration of ACh-producing neurons. |
| Glutamate | Major role mediating excitatory signals; involved in cognition, memory, and learning | Imbalance associated with memory loss in Alzheimer's disease; oversupply associated with neuron death and decreasing motor control in ALS patients. |
| GABA | Major role mediating inhibitory signals; involved in motor control and vision | Deficiency associated with seizures, tremors, and anxiety. |
| Norepinephrine | Major role controlling alertness and arousal; increases heart rate | Imbalance associated with mood disorders, manic depression. |
| Dopamine | Major role in award-motivated behavior; influences attention, emotion, and movement | Undersupply associated with tremors and muscle rigidity in Parkinson's disease; oversupply associated with schizophrenia and impulsive behavior. |
| Serotonin | Major role in regulation of mood, appetite, arousal, and sleep | Undersupply associated with depression, sleep problems, and compulsive behavior. |
| Nitrous Oxide | Involved in redox signaling; relaxes GI smooth muscle; vasodilator | Deficiency associated with depressed mood and impulsive behavior. |
| Endorphin | Pain and stress inhibitor | Undersupply associated with hypersensitivity to pain and depression; oversupply associated with insensitivity to pain and addictive behavior. |

the **thalamus** receives sensory information pertaining to each of the senses except for smell. The thalamus also receives response information from the cortex and directs it to the cerebellum and medulla. The **hypothalamus** is (unsurprisingly) located beneath the thalamus and is an important regulator of various motivated behaviors including hunger, thirst, sleep, body temperature, and sexual behavior. Together, the myelencephalon, metencephalon, mesencephalon, and diencephalon make up the brain stem. The **telencephalon** is the largest and most massive division of the human brain. Various fissures divide each hemisphere into four lobes:

1. The **frontal lobe** is involved with decision-making, planning, emotions, behavioral control, and motor control. Within the posterior portion of the frontal lobe, the **motor cortex** controls the planning and execution of voluntary movements.
2. Located at the top and rear of the head, the **parietal lobe** integrates sensory information, processes tactile sensory information, and helps us perceive the spatial location of objects and our own bodies. These functions are largely facilitated by the **somatosensory cortex**, a region of the parietal lobe that is adjacent to the motor cortex.
3. The **occipital lobe** is located at the back of the head and primarily deals with visual processing involving the visual cortex. Found within the most posterior portion of the occipital lobe is the visual cortex.
4. Located below the frontal and parietal lobes, the **temporal lobe** is involved with auditory processing, language comprehension, speech, and memory. Also located within the temporal lobes is the hippocampus, which plays a major role in the development and processing of memories, and the **amygdala**, which processes memories.

Fig. 6.1 depicts the lobes of the brain and selected structures of the cerebral cortex.

There are many techniques to study the brain:

1. The lesion method allows one to compare the subject's behavior pre-lesion and post-lesion to try to elucidate the functions of the targeted structure.
2. An **electroencephalogram (EEG)** measures the electrical activity emitted from the brain's neurons and produces an amplified readout. The resulting EEG reveals the gross electrical activity of the brain. Researchers often introduce

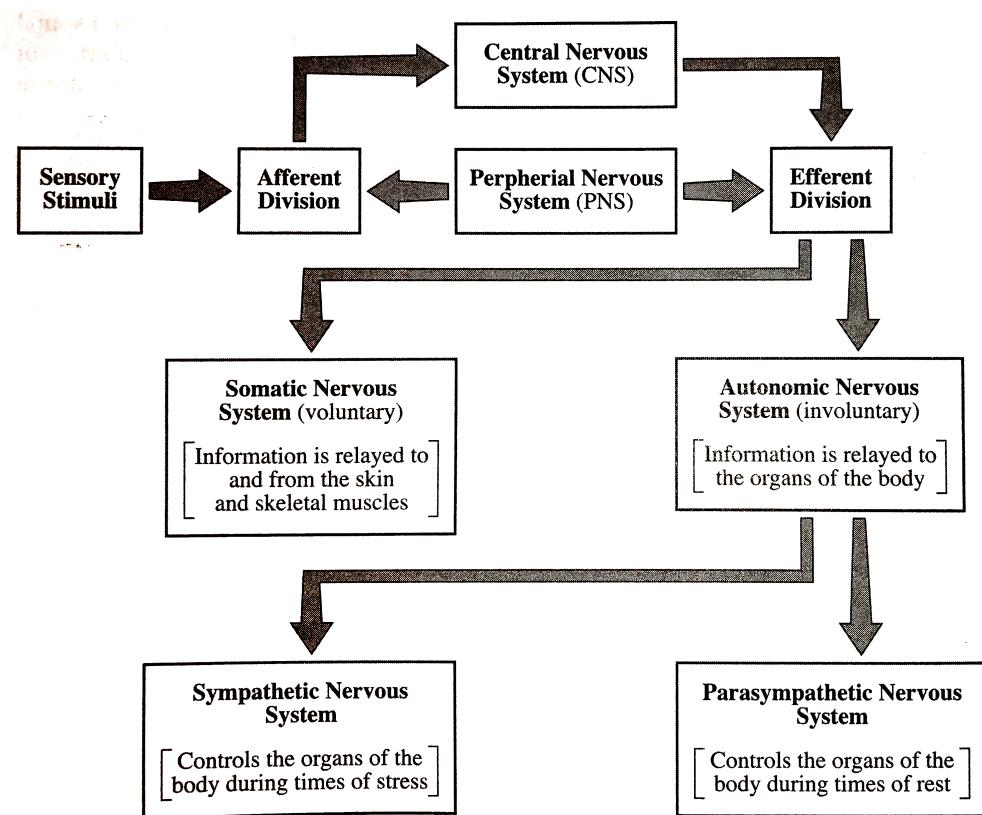


Figure 103: Functional divisions of the nervous system.

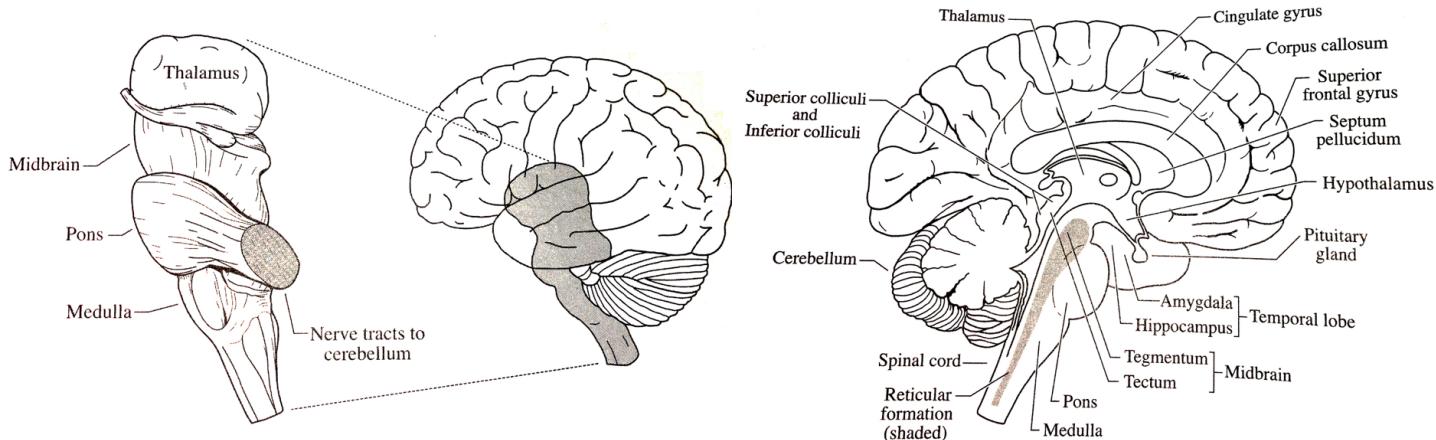


Figure 104: Major structures of the brain, including the components of the brainstem and the thalamus.

a specific stimulus to seek a particular response signal while having a computer filter out the unrelated noise. The resulting information is called the **sensory evoked potential** and is used to diagnose the cause of seizures clinically.

3. PET scans track each brain area's energy consumption. Radioactive 2-deoxyglucose (an analog of glucose) is injected into the carotid artery for direct delivery to the cerebral hemispheres. Because 2-deoxyglucose cannot be metabolized, it accumulates in active neuronal cells, allowing the PET-scan to identify 'hot spots' of neuronal activity while the

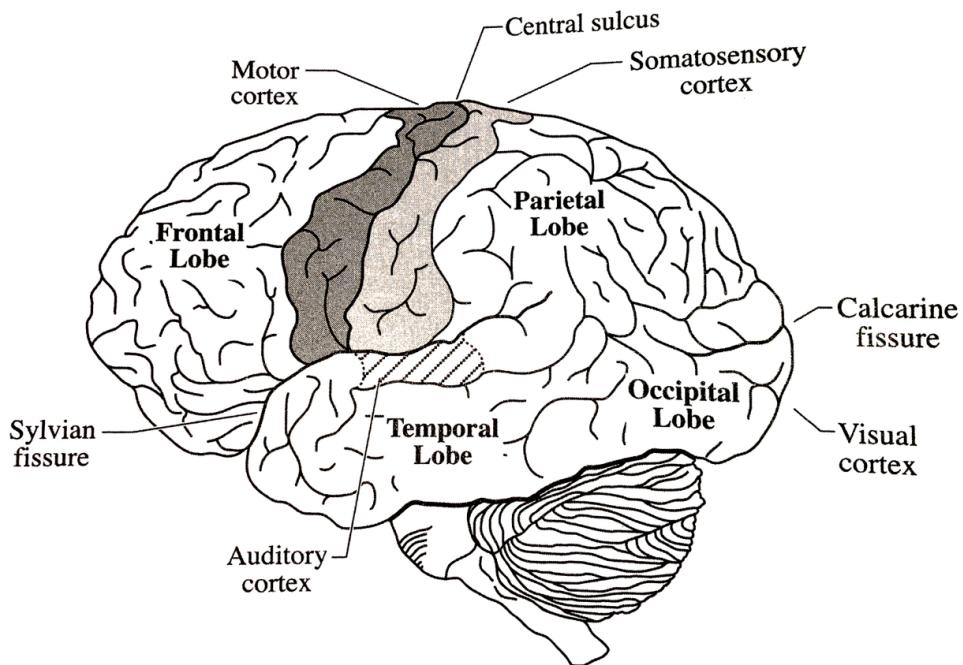


Figure 105: Lobes and selected structures of the cerebral cortex.

patient performs various tasks.

4. **Computed tomography (CT)** scans give a three-dimensional view of the brain using X-ray imaging.
5. **MRI** has higher spatial resolution than CT. There's also **functional MRI (fMRI)** that produces images of active brain areas by tracking the flow of oxygenated blood to these areas. This is possible because oxygenated blood has certain magnetic properties that alter the emitted hydrogen signals detected in a regular MRI scan.

Here are the major organs in the endocrine system:

1. The **pineal gland** influences the body's daily circadian rhythms.
2. The **hypothalamus** links the nervous system to the endocrine system by way of the pituitary gland (which it controls).
3. The **pituitary gland** secretes numerous peptide hormones, and is often referred to as the 'master gland' because of its significant influence over other organs and endocrine glands in the body.
4. The **thyroid gland** influences rate of cellular metabolism.
5. The **parathyroid glands** regulate blood calcium levels.
6. The **thymus** is an important component of the immune system.
7. The **adrenal glands** mediate the body's stress response, help maintain blood volume and pressure, and influence metabolism.
8. The **pancreas** regulates blood sugar levels.
9. The **ovaries** secrete female steroid sex hormones, and the **testes** secrete male steroid sex hormones.

and here are the major endocrine glands and the associated hormones that are released: **Behavior** is the way an animal adjusts to its environment. During development, here are the germ layer derivatives:

1. **Ectoderm:** Epidermis and associated structures, nervous system, sensory organs.

| Endocrine Gland | Hormones Released | Function |
|---------------------------|--|---|
| Hypothalamus | Thyrotropin-releasing hormone (TRH) Gonadotropin-releasing hormone (GnRH) Growth hormone-releasing hormone (GHRH) Corticotropin-releasing hormone (CRH) Somatostatin | Stimulates release of thyroid stimulating hormone (TSH) and prolactin from the anterior pituitary Stimulates release of follicle-stimulating hormone (FSH) and luteinizing hormone (LH) from the anterior pituitary Stimulates release of growth hormone (GH) from anterior pituitary Stimulates release of adrenocorticotrophic hormone (ACTH) from anterior pituitary Inhibits pituitary release of growth hormone (GH) and thyroid stimulating hormone (TSH) |
| Anterior Pituitary Gland | Growth Hormone (GH) Prolactin Adrenocorticotrophic Hormone (ACTH) Thyroid Stimulating Hormone (TSH) Luteinizing Hormone (LH) Follicle Stimulating Hormone (FSH) | Promotes growth in childhood and maintenance of healthy bone and muscle mass in adulthood Stimulates milk production in women Stimulates release of corticosteroids from adrenal glands Stimulates release of triiodothyronine h(T3) and thyroxine (T4) from the thyroid gland Acts on ovaries of females to regulate estrogen; acts on testes of males to regulate testosterone Simulates release of eggs in females and regulates normal sperm production in males |
| Posterior Pituitary Gland | Oxytocin Antidiuretic Hormone (ADH) (vapopressin) | Regulates female reproductive functions of childbirth and breastfeeding Regulates body's water balance |
| Thyroid Gland | Triiodothyronine (T3) Thyroxine (T4) | Important regulator of metabolism Important regulator of metabolism |
| Parathyroids | Parathyroid Hormone (PTH) | Regulates the body's calcium and phosphate levels to a very narrow range |
| Adrenal Cortex | Cortisol ('Stress Hormone') Aldosterone | Increases blood glucose levels; suppresses immune system Regulates body's balance of salt and water; helps regulate blood pressure |
| Adrenal Medulla | Epinephrine Norepinephrine | Increases heart rate and blood flow to muscles and brain; increases blood glucose levels Increases heart rate and alertness; acts as a vasoconstrictor |
| Pancreas | Insulin Glucagon | Decreases blood glucose levels Increases blood glucose levels |

2. **Mesoderm:** Bones, cartilage, circulatory system, dermis, excretory system, muscles, reproductive system.

3. **Endoderm:** Endothelium, lining of digestive tract, lining of respiratory system.

A girl's first menstrual period is called a **menarche** and usually arrives around age 12, whereas a boy's first ejaculation, called a **spermarche**, usually arrives around age 14 as a nocturnal emission. Changes to an adolescent's brain also result in significant cognitive development. With the onset of puberty, the brain ceases to make new synaptic connections and instead begins a process of **selective pruning** in which unused neurons and connections are eliminated. Adolescence also increases the growth of myelin, leading to improvements in judgment, long-term planning, and impulse control.

6.2 The Human Senses

Sensation refers to the passive process of detecting environmental stimuli through our various senses. **Perception** refers to the active process of receiving, reorganizing, and interpreting this raw, sensory information. We are only able to sense a stimulus when it reaches an *absolute threshold*—the point at which energy from a stimulus becomes detectable. This prevents us from feeling the trillions of microbes crawling on our skin or seeing ultraviolet rays. Beyond the absolute threshold, we can detect changes in the strength of a stimulus only if it exceeds our **difference threshold**. **Weber's Law** states that the magnitude of a just noticeable difference (JND) is proportional to the intensity of the original stimulus. This is why adding 2 more pounds to a 2-pound weight is noticeable, but adding 2 more pounds to a 200-pound weight is not. **Sensory adaptation** is the decrease in sensitivity to an unchanging stimulus as a consequence of constant exposure to that stimulus, which explains why we don't 'hear' the tick of a clock after prolonged exposure.

A neural receptor is classified by how quickly it adapts, or responds, to an incoming stimulus. **Phasic (rapidly-adapting) receptors** respond quickly and entirely to a stimulus, but stop responding even if the stimulus remains constant. **Tonic (slowly-adapting) receptors** respond gradually to a stimulus and provide a continuous signal for the entire duration of the stimulus.

Based on location, there are three types of receptors. **Exteroceptors** are located close to the body's surface and are specialized to detect external stimuli, such as tactile, gustatory, visual, olfactory, and auditory stimuli. **Interoceptors** are located within internal organs and are specialized to detect sensory information concerning the body's internal environment, such as blood pressure, pH levels, and dissolved gas concentrations. **Proprioceptors** are located in joints, muscles, tendons, and vestibular structures of the inner ear. They detect specific movement and positioning required for spatial reasoning and balance.

There are five primary types of receptors according to stimulus modality. **Chemoreceptors** detect chemical stimuli and play a primary role in olfaction and gustation. **Mechanoreceptors** detect stimuli that physically deform the tissue in which it resides, such as pressure or vibration. **Nociceptors** are sensitive to damaging and painful stimuli and are almost always tonic. **Photoreceptors** respond to EMR (photos) and are essential for sight. **Thermoreceptors** detect warmth and cold and are almost always phasic.

A **sensory pathway** describes the stimulus-induced relay of information to the CNS for complex integration and perception. Generally speaking, there are three different neurons involved. The **primary neuron** is the afferent/sensory neuron and sends the signal to the spinal cord. The **secondary neuron** then picks up the signal and sends it to the thalamus for processing and redirection to the correct sensory area of the cerebral cortex. The redirection is performed by the **tertiary neuron**. All sensory inputs follow this mechanism except for **olfaction**, which instead relays the sense stimulus to the **olfactory bulb**, then relaying it to the primary olfactory cortex in the temporal lobe, and then the limbic system in the frontal cortex.

Motion parallax is the perception of fast-moving objects to be closer than slow-moving objects. This is why when we're riding in a car, the road appears to whizz by while the mountains in the distance appear to move much slower. Another important aspect of motion perception is the **phi phenomenon**—apparent motion caused by the rapid succession of stationary stimuli. For example, when you watch a movie, you are seeing many pictures flash in rapid sequence, not actual movement. Due to the phi phenomenon, your brain connects these pictures to create the perception of movement.

Perception is also organized in such a way as to maintain perceptual stability under circumstances of changing stimuli. For example, imagine that your friend is wearing a bright red shirt. Now, when your friend steps into the shade, the shirt is a much darker red. If you constantly had to reprocess incoming sensory information every time an object changed, you would be amazed by your friend's ability to change the color of her shirt instantaneously. But this doesn't happen thanks to **perceptual constancy**, the tendency to recognize familiar objects as having the same properties despite changes in their lighting (i.e. color constancy), distance (i.e. size constancy), or angle of perspective (i.e. shape constancy).

Sensory information can be organized into groups according to the **gestalt principles**, which demonstrate how the sum of sensory parts actually differs from the perceived whole. Fig. 6.2 shows examples of the four primary gestalt principles: similarity, proximity, continuity, and closure.

Vision: Light enters the eye through the **cornea**, a transparent and protective membrane that functions as an outer

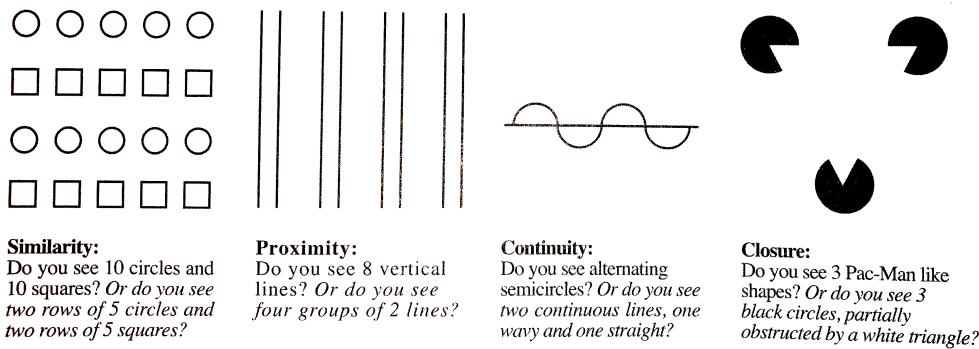


Figure 106: The gestalt principles. The gestalt principles predict that you will see the italicized result.

lens, bending light to provide roughly 70% of the eye's total focusing power. The cornea doesn't have blood vessels in order to be transparent; its nutrients are acquired through the diffusion of tear fluid. From the cornea, light travels through the **pupil**, an opening of adjustable diameter that is regulated by surrounding contractile tissue known as the **iris**. This colored muscle constricts or dilates to maintain a balance between *visual acuity* and *visual sensitivity*. Immediately behind the pupil is a **lens** that focuses incoming light onto the **retina**, a multi-layered, light-sensitive tissue that lines the inner surface of the eye. **Ciliary muscles** act on the suspensory ligaments holding the lens in place to apply adjustable increments of tension necessary for focus. When we focus on distant objects, the ciliary muscles relax, causing the suspensory ligaments to become taut, which flattens the lens, effectively increasing the *focal distance*. When we focus on nearby objects, the ciliary muscles contract, causing the suspensory ligaments to contract, which curves the lens, gaining increased ability to refract light.

There are other important structures that do not alter light's path to the retina:

1. The **sclera** is a fibrous, protective layer that covers the eye and provides rigidity. It's the visible, white portion of the eye.
2. The **aqueous humor** is located in the *anterior chamber* (space between the cornea and the lens), and is a gelatinous fluid that maintains the intraocular pressure of the eye, supplies nutrients, and removes debris.
3. The **vitreous humor** is a gelatinous fluid found between the lens in the *vitreous chamber*. It maintains the eye's shape while also firmly keeping the retina in place by pressing it to the **choroid**, a layer of connective tissue and blood vessels that nourishes the posterior structures of the eye.
4. The **extraocular muscles** control gross eye movement and eyelid elevation.

Fig. 6.2 highlights the anatomy of the eye.

Fig. 6.2 shows the process of **phototransduction** and the involved cell types. The **amacrine cells** and the **horizontal cells** synapse across the other neurons to communicate laterally and coordinate sensory input. When light activates **photoreceptor cells**, a biochemical cascade results in decreased *glutamate* release, which in turn affects the activity of the **bipolar cells**, which in turn relays this inhibitory or excitatory response to retinal ganglion cells. The axons of **retinal ganglion cells** exit the retina via the **optic nerve** and carry visual sensory information to the brain.

It is important to note that the retinal layers are, in a sense, backwards. Light does not immediately hit the inner retina, undergo phototransduction, and create a neural impulse that exits the outer retina in a linear fashion. Instead, light must pass through four neuronal layers before activating the photoreceptors to generate a neural message. This information is then passed back in the opposite direction to the retinal ganglion cells located across the innermost layer of the retina.

There are two types of photoreceptor cells: cones and rods. **Cones** are responsible for color vision and **rods** are responsible for peripheral vision, contrast and edge perception, and nighttime vision. The differences and similarities of cones and rods are shown in Fig. 6.2.

When light reaches the photoreceptor layer, it strikes either a cone or a rod to initiate the phototransduction signal transduction pathway, which relies on rhodopsin in rods and photopsin in cones, and cGMP as the secondary messenger. As a result, less glutamate neurotransmitter is released onto bipolar cells, which process this fact as a light signal.

The primary visual pathway is the **retina-geniculate-striate pathway**, which conducts retinal signals through the

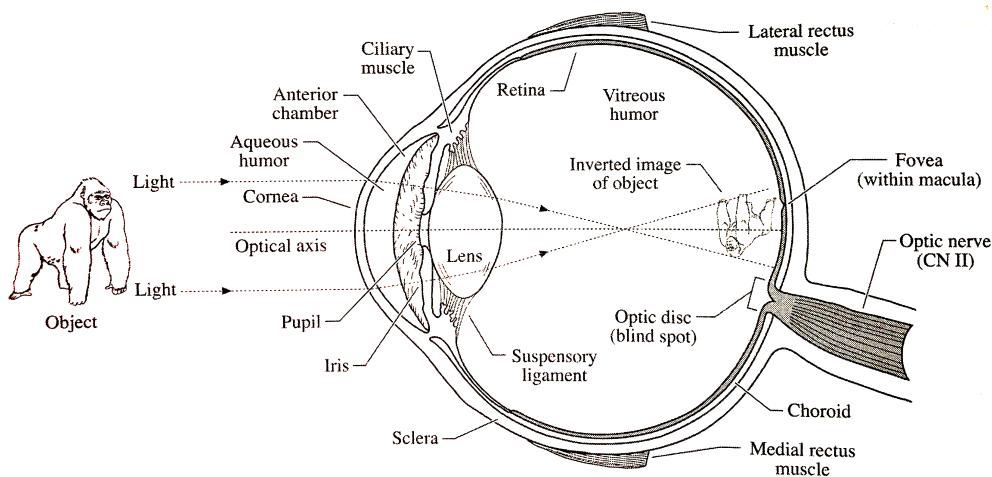


Figure 107: The major components of the eye.

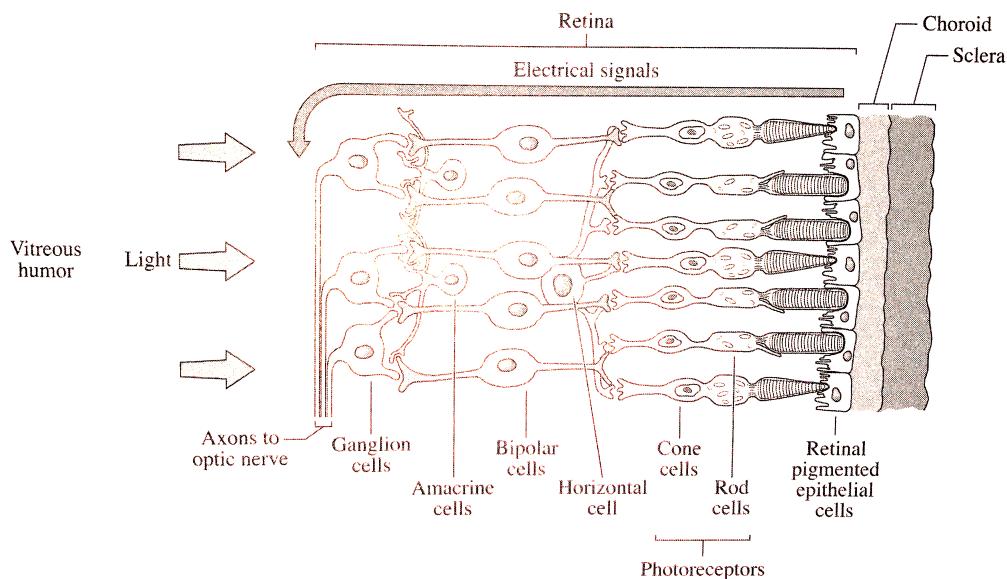


Figure 108: Directional flow of light and electrical information through the layers of the retina.

lateral geniculate nuclei (LGN) of the thalamus to the **primary visual cortex** located in the posterior occipital lobe. The outer portion of the retina, known as the **temporal hemiretina**, receives visual signals from the opposite visual field and sends them ipsilaterally (along the same side) to the primary visual cortex. The inner portion of the retina, known as the **nasal hemiretina**, receives visual signals from its like-sided visual field and sends them contralaterally to the opposite primary visual cortex via the **optic chiasm**. This is further shown in Fig. 6.2. The retina-gangetic-striate pathway is further divided into two separate sub-pathways: the *parvocellular pathway* and the *magnocellular pathway*. Parvocellular neurons specialize in the *spatial resolution* of an object; they primarily receive input from cones and are sensitive to color, stationary objects, and fine pattern details. The magnocellular neurons specialize in the *temporal resolution* of an object; they receive input from rods and are particularly sensitive to movement.

Audition: The ear is organized into three sections: the outer, middle, and inner ear. Sound waves enter the outer ear through the **pinna**, which is the visible, curved piece of cartilage that 'catches' sound. From the pinna, sound waves become amplified in the **auditory canal**.

Now in the **middle ear**, incoming sound waves cause the eardrum to vibrate, which in turn vibrates three small bones

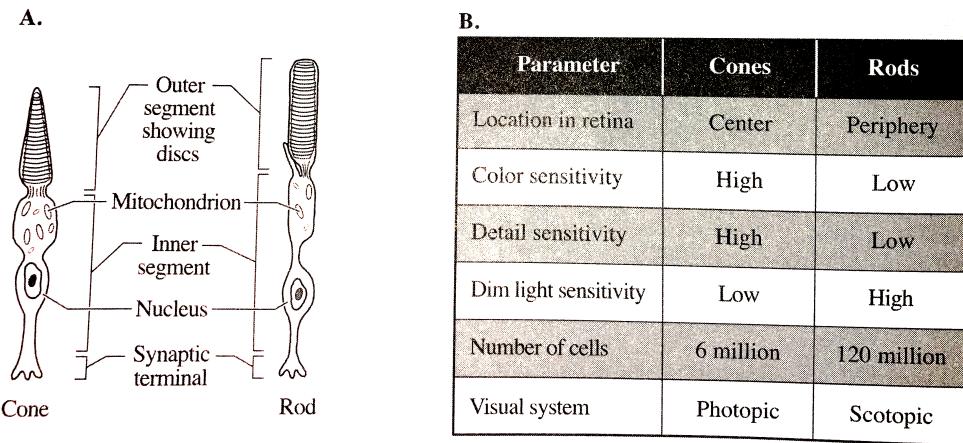


Figure 109: The major components of and differences between a cone and a rod.

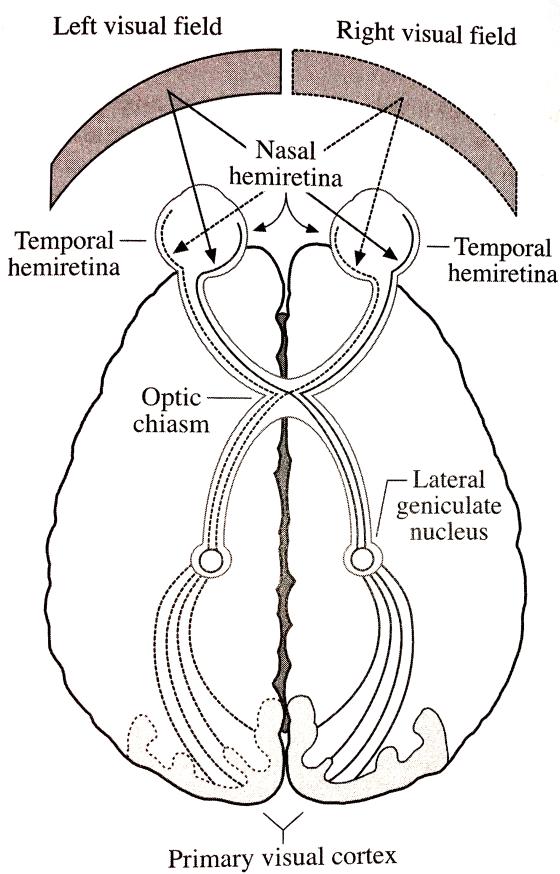


Figure 110: The retina-geniculate-striate pathway.

known as **ossicles**. Connected to the **tympanic membrane** (another word for the eardrum), the **malleus** detects these vibrations and relays them to the **incus**, which in turn vibrates the **stapes**. The main function of the ossicles is to amplify sound vibrations. Prolonged exposure to loud sound triggers the **stapedius muscle**, the muscle connecting the eardrum and small bones, to contract and pull the tympanic membrane and ossicles in opposite directions. This causes the tympanic

membrane to become more rigid and less sensitive to vibration, effectively dampening the noise and protecting the inner ear.

Vibrations from the stapes reach the **oval window**, a thin membrane connecting the middle ear to the **inner ear**, causing it to vibrate the fluid within the **cochlea**. Because fluids are fairly incompressible, pressure needs to be relieved as the oval window moves in and out. To accomplish this, a structure at the other end of the cochlea, called the **round window**, will move in the direction opposite to the oval window. The cochlea is a long, snail-shaped tube that contains an internal membrane called the **organ of Corti**. The organ of Corti partitions the cochlea into a longer, narrower tube and is composed of two membranes: the **basilar membrane** and the **tectorial membrane**. The specialized receptors of the auditory system, **hair cells**, are a type of mechanoreceptor mounted in the basilar membrane. They convert pressure differences into neural signals that are sent to the brain via the auditory nerve. **Fig. 6.2** highlights the location of these structures.

Auditory transduction occurs through the hair cells of the inner ear. Influx of potassium is induced mechanically, and

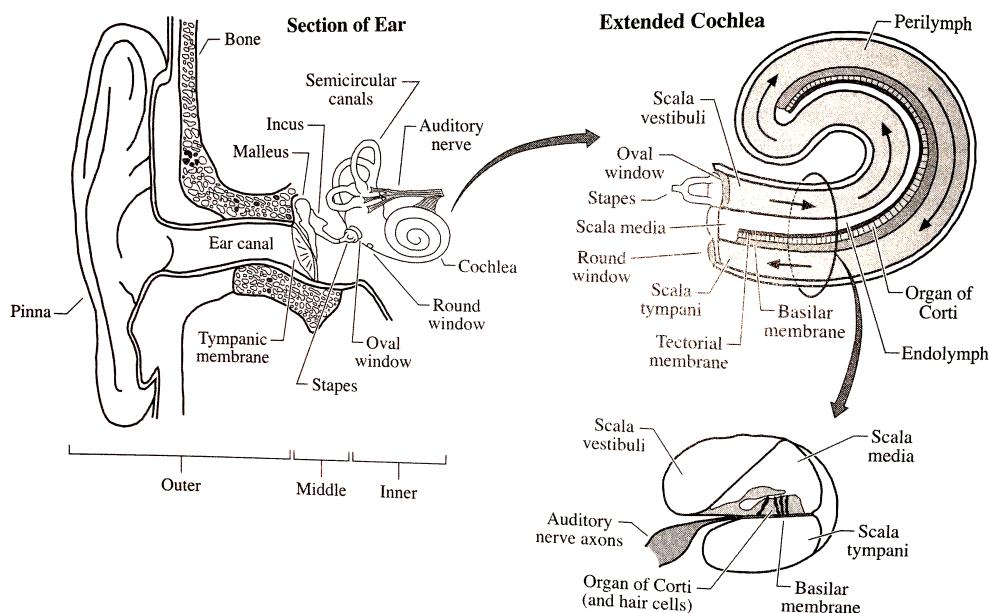


Figure 111: The major anatomical structures of the ear.

is caused by a rapid influx of potassium ions from the surrounding **endolymph**, a unique fluid with an unusually high potassium concentration and unusually low sodium concentration. The influx leads to depolarization of the hair cell.

The axons of each auditory nerve, one exiting from each cochlea, synapse ipsilaterally to collections of neurons in the hindbrain called the **cochlear nuclei**. From here, information may *decussate* (cross from one side of the brain to the other) at the **trapezoid bodies** or travel ipsilaterally to synapse at the **superior olivary complex**. Axons from olivary neurons then travel through the **lateral lemniscus** and synapse to neurons in the **inferior culiculi** of the tectum. From here, information travels to the **medial geniculate nuclei** of the thalamus, which in turn relays this information to the primary auditory cortex for processing.

The **semicircular canals** attached to the cochlea and the **otolith organs** are the receptive organs of the **vestibular system**, the sensory system responsible for providing one's awareness of head rotation, gravitation, and movement. The three semicircular canals specialize in detecting head rotation. The otolith organs include two sac-like structures located between the semicircular canals and the cochlea, called the *utricle* and *saccule*, which specialize in detecting horizontal and vertical acceleration.

Somatosensation: Each cutaneous (skin) receptor has the same general structure: a neuron with its cell body located in the *dorsal root ganglion* of the spinal cord and dendrites that are sensitive to specific stimuli. The variation between cutaneous receptors comes from the differences between these dendrites. *Meissner's corpuscles* and *Pacinian corpuscles* are classified as *fast-adapting receptors*, whereas *Merkel's discs* and *Ruffini endings* are classified as *slow-adapting receptors*. Unlike these four encapsulated cutaneous receptors that are sensitive to pressure, stretch, and vibration, **free nerve endings** are simply exposed dendrites and are particularly sensitive to temperature and pain. Meissner's corpuscles and Merkel's discs are specialized to detect pressure within a small receptive field; Ruffini's endings are specialized to detect stretch within a large receptive field; Pacinian corpuscles are specialized to detect vibration within a large receptive field. The size of each

receptor's receptive field is inversely correlated to its spatial resolution—for example, a Pacinian corpuscle is able to detect a vibrating stimulus over a relatively large area, but is able to provide only very basic information regarding its border and location. All these receptors generally rely on mechanosensitive ion channels; when activated mechanically, they allow ions to enter the cell and stimulate an action potential.

Temperature and pain are carried along two different nerve givers: A δ fibers and C fibers. **A δ fibers** possess myelinated axons and conduct neural signals at a moderate rate. **C fibers** include unmyelinated axons and conduct neural signals relatively slowly. Consider the sensation of burning yourself by touching a hot stove by accident. Carried by A δ fibers, fast pain makes you recoil from the stovetop even before you feel burned. Then, slow pain carried by C fibers brings the duller yet persistent burning sensation some seconds afterwards.

Somatosensory information can be organized by location. Distinct areas of skin, called **dermatomes**, are innervated by individual peripheral nerve fibers that enter the spinal cord at different points. Input from different dermatomes remains segregated along the various pathways to the somatosensory cortex, and is known as **somatotopic organization**, allowing the brain to create a spatial map of the body represented in the somatosensory cortex. This ‘map’ is commonly referred to as the **somatosensory homunculus**.

Olfaction and Gustation: The olfactory receptors are embedded in the **olfactory mucosa**, a mucus-covered tissue located in the upper portion of the nasal cavity that allows odorants to dissolve and become more readily detected by olfactory receptor dendrites. The axons of olfactory receptors travel to the **olfactory bulbs** where they synapse with secondary neurons called **mitral cells** that carry information to the brain. From there, information travels ipsilaterally to the **primary olfactory cortex** of the medial temporal lobes for processing. The olfactory system is unique among all senses, because information about smell is not relayed through the thalamus before reaching the cortex, as is the case for the other major senses. After reaching the primary olfactory cortex, information travels to the structures of the **limbic system** (i.e. hypothalamus, amygdala, and hippocampus), for emotional responses and memory formation, and to the **orbitofrontal cortex**, located in the anterior portion of the frontal lobes, for conscious perception.

The axons of Olfactory transduction relies on GPCRs—the activated G-protein will produce cAMP, which in turn cause an influx of Na $^+$ and Ca $^{2+}$ ions, leading to cell depolarization and the generation of an action potential.

Gustatory transduction is triggered by either a chemical’s direct passage through an ion channel or a G-protein coupled receptor. For example, ion channels mediate salty and sour tastes, whereas GPCRs mediate sweet, bitter, and umami tastes.

1. **Salt Receptors:** The influx of sodium ions such to Na $^+$ ion channels depolarizes the cell, causing voltage-gated calcium channels to open and the generation of an action potential.
2. **Sour Receptors:** We have H $^+$ ion channels, proton-gated ion channels that either close when bound to H $^+$ to prevent K $^+$ ions from leaving the cell or open when bound to H $^+$ to allow Na $^+$ ions to enter the cell. This increases the influx of H $^+$ and Na $^+$ and decreases the efflux of K $^+$ to generate an action potential.
3. **Sweet Receptors:** Natural sugars and related chemicals bind to GPCRs that cause these receptors to release the **gustducin** G-protein, which in turn reduces the efflux of potassium ions and hence the generation of an action potential.
4. **Bitter Receptors:** These also have the exact same receptors and mechanism of sweet receptors, but have the additional step of using gustducin to stimulate the endoplasmic reticulum to release the Ca $^{2+}$ ions stored in the lumen, further helping the generation of an action potential.
5. **Umami Receptors:** These respond primarily to *glutamate*, amongst other amino acids, via the same GPCR cascade described for bitter receptors.

Neural information from each receptor is transmitted ipsilaterally via the **facial nerve** and **glossopharyngeal nerve** to the **solitary nucleus** of the medulla. From here, secondary neurons terminate in the **ventral posteromedial nucleus (VPM)** of the thalamus, where they synapse with tertiary neurons that transmit information ipsilaterally to the **primary gustatory cortex**, located near the postcentral gyrus in the parietal lobe.

6.3 Learning and Memory

Learning is the process of acquiring new and relatively permanent information or behaviors. Before we discuss learning, we need to talk about stimuli. A **stimulus** is any event or situation that evokes a response. There are three types of stimuli: neutral, conditioned, and unconditioned.

1. A **neutral stimulus (NS)** is a stimulus that initially produces no response other than focusing attention, such as a tone, click, light, etc.
2. An **unconditioned stimulus (US)** is a stimulus that naturally and automatically triggers a response; its effects do not depend on previous experiences. Common examples are food, water, sucrose, shocks, and temperature changes.
3. A **conditioned stimulus (CS)** is a stimulus that was previously neutral, but after becoming associated with the unconditional stimulus, it eventually elicits a conditioned response.

An **unconditioned response (UR)** is a response that is unlearned and occurs naturally with presentation of an unconditioned stimulus. A **conditioned response (CR)** is a learned response to a previously neutral stimulus. Now, there are three types of learning.

Nonassociative learning is a change in response to a stimulus due to repeated exposure to that stimulus. The two forms of nonassociative learning are *habituation* and *sensitization*. **Habituation** is a process in which the magnitude of a response to a specific stimulus decreases with repeated exposure to that stimulus. As the frequency of the stimulus increases, the rate of habituation can increase, but if the *strength* of the stimulus increases, habituation is slowed or even stopped (this makes sense). Following habituation, a single presentation of a different stimulus restores the response to the original stimulus. This response recovery is called **dishabituation** and results in a renewed interest. In other words, habituation is reversible. **Sensitization** is the process in which the magnitude of a response to a stimulus increases with repeated exposure to that stimulus.

Associative learning is learning that one event occurs with another. The two types of associative learning are classical conditioning and operant conditioning. **Conditioning** is a general term defined as the process in which a particular behavior becomes dependent upon the presence of a stimulus. In **classical conditioning**, learning occurs when a conditioned stimulus is paired with an unconditioned stimulus, causing an association to form between the two. An example of this is Pavlov's experiment where he trained a dog to salivate in response to a bell. The presence of food is the *unconditioned stimulus*, the act of salivating is the *unconditioned response* that became a *conditioned response*, and the bell is the *neutral stimulus* that became a *conditioned stimulus*.

Acquisition is the initial learning of the *stimulus-response relationship*, which is the idea that a stimulus automatically triggers a response. The time interval between CS and US, the salience and sensory modality of the CS, and the order of CS/US presentation can all affect the rate of acquisition:

1. **Forward conditioning** is where the presentation of the CS precedes the presentation of the US, and usually produces the quickest learning. This is the most biologically relevant, as if the predictive cue followed the good or bad event, there would be no time preparation. In *delay conditioning*, the CS is presented before the US and continues throughout the presentation of the US. In *trace conditioning*, the CS is presented and then the US is presented with no overlap.
2. **Simultaneous conditioning** is where the CS and US are presented and terminated simultaneously.
3. **Higher-order conditioning** (also known as *second-order conditioning*) is a procedure in which an existing CS is paired with a NS, creating a second (often weaker) S. For example, if a tone is associated with a shock and forward conditioning is performed to associate a light with the tone, the light alone will come to elicit the CR of a jump.
4. **Backward conditioning** is where the US is presented before the CS (e.g. the food is presented before the bell).

Extinction is the diminishing of a conditioned response due to repeated exposure of the CS in the absence of the US. For example, in Pavlov's classic experiment, after the tone was associated with salivation, he repeatedly presented the tone (CS) in the absence of food (US), and eventually the dog salivated less and less with each subsequent tone. **Spontaneous recovery** is the reappearance of a weakened conditioned response after a rest period. The occurrence of spontaneous recovery suggests that the CR is being suppressed rather than unlearned during extinction. This is shown in **Fig. 6.3**.

Stimulus generalization occurs when a conditioned response to a specific stimulus transfers to other stimuli similar to the CS. For example, if I am scared of rats, I could also become scared of a white rabbit or Santa Claus's beard since they are all white furry objects. The counterpart to this is **stimulus discrimination**.

The second type of associative learning is called **operant conditioning**, where learning occurs by associating an action with a consequence. These learned associations allow an organism to act on the environment to maximize reward and minimize distress. Through operant conditioning, a rat can be taught to press a lever, a pigeon to peck a light, a raccoon to drop a coin into a coin, etc. The procedure by which such behavior is learned is called **shaping**. Rewarding specific behaviors on a step-by-step basis gradually to build a response can shape complex behavior.

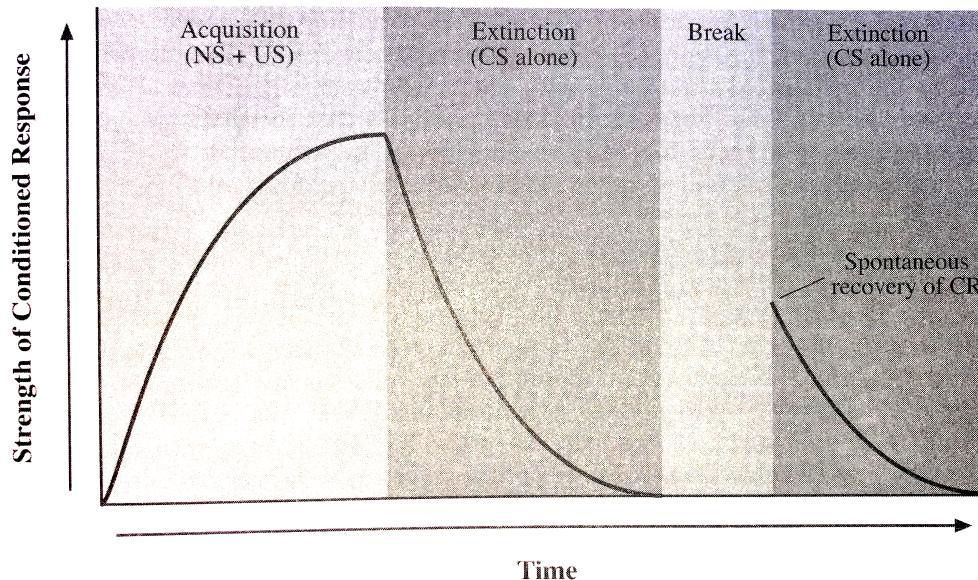


Figure 112: Acquisition, extinction, and spontaneous recovery.

In operant conditioning, **extinction** is the reduction of an instrumental behavior when it is no longer reinforced. While a **discriminative stimulus** is a stimulus that predicts the presence of reinforcement, an **extinction** predicts the absence of reinforcement. For example, if a lever press earns food only when a tone is playing, the tone becomes a discriminative stimulus. But, if a rat learns that a lever press does not earn food while a click is playing, the click becomes an extinction stimulus.

Positive reinforcement strengthens a response by presenting a pleasurable event or outcome after the response, while **negative reinforcement** strengthens a response by removing or reducing an unfavorable event or outcome. **Escape behavior** is a response that *terminates* an unpleasant stimulus that is already present. For example, before I had bad smell, so my behavior is to take a shower so that the bad smell is gone. **Avoidance behavior** is a response that *prevents* an unpleasant stimulus. For example, before I had no bad smell, but my behavior is to apply deodorant so that I can prevent the bad smell.

Positive punishment weakens a response by presenting an aversive stimulus after the response. For example, giving a toddler a spanking for throwing a tantrum or receiving a bad grade. **Negative punishment** weakens a response by removing a desirable stimulus following the response. For example, taking away a teenager's cell phone for swearing is negative punishment.

Surprise is defined to be a difference in what is expected and what is observed. In order to form a CS-US association, there must be some degree of surprise. The expected could be as simple as 'nothing should happen,' while the observed may be 'food is presented in front of me.'

Observational learning is a form of social learning that introduces the influential effects of a model, someone of authority whose own behavior changes the behavior of another. This is especially important during childhood. Observational learning is based in **mirror neurons**, which are a type of neuron that fires both when an animal performs a specific behavior and when the animal observes another performing that specific behavior. The mirror-neuron system within the brain is responsible for making the observer feel what the observed action must feel like, and is therefore essential in social interactions.

Memory is the persistence of learning over time through the encoding, storage, and retrieval of information. **Encoding** is the process of transforming information into a usable construct that can be stored within the brain. The process begins with attention to sensory input, which then becomes encoded within the various sensory areas of the cortex. The perceived sensations are then combined into a single construct in the *hippocampus*, which acts as the brain's sorting center, deciding which sensory inputs will eventually become stored memories. **Synaptic plasticity** is the ability of neuronal synapses to strengthen or weaken in response to increases or decreases in their activity. There are four types of encoding:

1. **Visual encoding** is the process of transforming visual sensory information. Sensory input such as color, shape, or size that is visually perceived is temporarily stored within our **iconic memory**.

2. **Acoustic encoding** is the process of transforming auditory sensory information. Auditory input such as words, melodies, and noises is temporarily stored within our **echoic memory**.
3. **Tactile encoding** is the process of encoding the sense of touch. Sensory input including vibration, texture, and pressure is temporarily encoded within our **haptic memory**.
4. **Semantic encoding** is different from the other three in that meaningful or contextual sensory information is encoded, as opposed to neutral sensory input. Information stored using semantic encoding typically is stronger and easier to retrieve because it has been associated with something that already has meaning.

The process of encoding can also happen either automatically or effortfully. Different effortful memory strategies include chunking, mnemonics, and acronyms. Chunking is basically like memorizing the number 1776200119692009 as the four years 1776, 2001, 1969, 2009. **Distributed practice** is a learning strategy that aids memory formation by distributing study or practice over time for better encoding. The more strongly a memory is encoded, the easier it is to retrieve in the future.

Once sensory information is encoded as a usable construct, it must be stored for later retrieval. **Storage** is the process of retaining information. Where the information is stored, the capacity that can be stored, the duration of a memory, and the kind of information retained all characterize the different aspects of memory storage.

The earliest stage of memory that records sensory information is called **sensory memory**. After a scene is perceived, the sensory memory holds information regarding what is visually sensed for less than one second. Previously, we introduced three types of sensory memory: iconic, echoic, and haptic. Each of these systems holds sensory information on the order of seconds and decays quickly. Only a fraction of the information that enters these systems passes into the next stage of short-term memory. **Short-term memory (STM)** also stores information temporarily, on the order of seconds and minutes, before it is lost. **Working memory** is STM's underlying support system that processes, manipulates, and stores current information. While STM is static, working memory is dynamic. The working memory has four subsystems:

1. The **phonological loop** is the subsystem dedicated to the temporary storage of auditory information.
2. The **visuospatial sketchpad** is the subsystem dedicated to the temporary storage of visual information.
3. The **episodic buffer** serves to integrate representations from the phonological loop and the visuospatial sketchpad in an ordered, timely, and coherent sequence.
4. The **central executive** is responsible for controlling and coordinating the activities of the other three memory systems, ensuring that attention and resources are allocated properly.

Since short-term memory has a limited capacity, information that is not attended to can be easily forgotten. **Rehearsal** refers to the mental techniques that strengthen the retention of memories. There are two types of rehearsal: *maintenance rehearsal* and *elaborative rehearsal*. In **maintanence rehearsal**, the target piece of information is continuously repeated. For example, in order to remember a phone number long enough to dial it, people have a tendency to repeat the information over and over. In **elaborative rehearsal**, the meaning of the target piece of information is considered and further associated with previously stored information.

Long-term memory (LTM) is the third and last stage of the three-stage memory model. LTM stores a relatively limitless amount of information for a relatively indefinite amount of time. The more attention that is given to a memory trace in short-term memory, the greater the strength of an engram within long-term memory. **Consolidation** is the process of stabilizing this memory. **Reconsolidation** is the process of recalling previously stored information and consolidating it again in order to maintain, strengthen, or modify a memory. Once a memory is consolidated, it is considered to be in long-term storage and stable. However, this does not imply the memory cannot be strengthened, weakened, or modified. Reconsolidation retrieves a stored memory and places it into one's working memory—in turn, making the memory **labile**, or readily able to change.

Retrieval is the process of accessing stored memories. The idea that external cues can ease one's ability to remember something is demonstrated through contextual learning. Have you ever walked upstairs to get something, only to realize that you can't remember what it was? Then, as you walk downstairs back into your living room, the memory seems to return suddenly? This phenomenon can be attributed to **context-dependent memory**. Since the retrieval of memories is highly influenced by retrieval cues, returning to the thought's original context can help jolt your memory. Memory is also **state-dependent**. For example, if you're happy, the state of being happy can activate the retrieval of other happy memories. There's also the **serial position effect**, which is the tendency to remember the first and last items in a list best. These are some of the processes that aid retrieval.

There are three types of memory retrieval. **Recall** is remembering information that was previously learned but is currently not physically present. Fill-in-the-blank questions on a test are examples that ask you to recall information. **Recognition** is identifying information that was previously learned after being cued. Multiple-choice questions are examples that require you to recognize the correct answer from what is presented. **Relearning** is the increased rate of learning information that has previously been learned.

There are six main reasons for why people forget things:

1. **Ineffective encoding** is when information from the working/short-term memory fails to ever enter long-term storage. For example, we don't *exactly* remember the pattern on a penny, because we never bothered to memorize it. For this reason, we are not specifically 'forgetting' the details of the penny, because we never stored the information in the first place.
2. The **decay theory** attributes the loss of memory entirely to the function of time.
3. The **interference theory** attributes forgetfulness to ineffective processes of retrieval, and suggests forgetting occurs due to memory disruptions from both previously learned information and newly presented information. There are two forms of interference: proactive and retroactive. In **proactive interference**, old memories disrupt the retrieval of newer memories. For example, if you are trying to learn Spanish, but previously learned French in high school, you might have difficulty remembering the new Spanish verb tenses and vocabulary. In **retroactive interference**, new memories disrupt the retrieval of older memories. An example of this would be if you find it difficult to recall the concept of molecular mass in chemistry, which you learned a week ago, because you learned the concept of density today.
4. **Cue-dependent forgetting/context-dependent forgetting** also attributes forgetfulness to ineffective processes of retrieval, and is based on the idea that the target information is present in long-term memory, but it's currently and temporarily inaccessible. In other words, if the right retrieval cue is presented, the memory can be made available.
5. **Motivated forgetting/repression** is a defense mechanism the mind employs to forget unpleasant memories. This theory proposes such forgetting as a mechanism to banish anxiety-associated memories into the unconscious in order to protect the image we have of ourselves.
6. **Amnesia** is forgetting due to 'organic causes,' such as physical trauma, neurological disease, and aging. There are two main types of amnesia: *anterograde* and *retrograde*. People suffering from **anterograde amnesia** have an inability to form new long-term memories; they are still able to recall childhood memories and any information stored prior to the organic cause. Conversely, people suffering from **retrograde amnesia** have an inability to retrieve old memories formed prior to a particular date, such as when the organic damage occurred.

There are two types of long-term memory: explicit and implicit memory. **Explicit memory**, also known as **declarative memory**, is information you know and can tell. There are two types of explicit memory: episodic memory and semantic memory. **Episodic** describes memory for specific events or experiences, while **semantic** describes memory for general knowledge and facts. The two main structures of the brain involved in processing explicit memories are the frontal lobes and the hippocampus. When a person retrieves explicit memories, the brain sends a collection of information to the *frontal lobes* to be processed by working memory.

Implicit memory, also known as **nondeclarative memory**, is information you know and can show. It is most often thought of as memory without awareness; it is both unconscious and unintentional. There are three types of implicit memory: priming, procedural, and classical conditioning.

Through **priming**, a person's future behavior is influenced by previous subconscious preparation. For example, if your best friend swears constantly and uses inappropriate language to talk about others, then when your mom asks about your new teacher, you are predisposed to describe him negatively. **Procedural memories** allow us to perform daily tasks, especially motor skills, without thought. Typing on a keyboard, playing an instrument, and using a pencil all rely on procedural memory. Through *classical conditioning*, implicit memories influence behavior based on previously learned associations.

The two main structures of the brain involved in processing implicit memories are the **cerebellum** and the **basal ganglia**. Emotion-related memory is also important—our emotions trigger hormonal changes in the body that influence memory processing. Most importantly, these hormones recruit the **amygdala** and give it a vital role in memory formation. Because of the amygdala, emotional arousal can result in stronger, more vivid memories.

Source amnesia is the inability to remember the source of previously learned information. For example, I know that all porcupines float in water, but I cannot recall whether I read this in a newspaper, heard it on television, or imagined

it in a dream. When memory loss begins to interfere severely with a person's daily functioning, it is called **dementia**. Another neurological disorder that results in memory dysfunctions is **Korsakoff's syndrome**. Korsakoff's syndrome is a brain disease characterized by a thiamine (Vitamin B1) deficiency, most often associated with alcohol abuse.

6.4 Emotion, Stress, Motivation, and Addiction

Emotions are temporary, subjective experiences based on your physiological arousal, your interpretation of the recent events, and your behavioral response. The emotions experienced are an interplay of physiological, cognitive, and behavioral functions. The *physiological component* refers to the body's physical changes in response to a situation. The *cognitive component* of emotion accounts for a person's interpretation of a situation. A person's expectations, thoughts, and beliefs can control how a particular situation is perceived, and thus experienced. The *behavioral component* includes the bodily expression of emotion, both observed in others and felt within one's own body. Nonverbal body language, such as posture and hand gestures as well as facial expressions, plays a significant role in the feeling of an emotion. There are three different theories of emotion: the James-Lange theory, the Cannon-Bard theory, and the Schachter-Singer theory:

1. The **James-Lange theory** of emotion posits that physiological arousal and behavioral response *precede* the experience of an emotion. In other words, seeing the gorilla jump out of the closet caused your heart rate to increase (physiological arousal) and you to run away (behavioral response), thereby instigating the emotion of fear. According to James-Lange, the emotion was secondary to the primary physiological changes and behavioral responses.
2. The **Cannon-Bard theory** of emotion proposes physiological response and conscious awareness of the emotion are both experienced simultaneously. In other words, after seeing the gorilla, you became consciously frightened and physiologically aroused at the same time. One component did not precede the other, but rather the two were concurrent. A salient stimulus in the environment activates the brain's relay station, the thalamus, in turn activating the autonomic nervous system to initiate bodily changes and activating the amygdala and cortex to initiate the emotion.
3. The **Schachter-Singer theory**, also known as the **two-factor theory**, states that cognitive evaluation of one's physiological arousal results in the experience of an emotion. This theory expands upon the previous two theories by providing *direction*. The idea that previously arousing events can affect one's reaction to the next event is called the **spillover effect**. For example, if you slept through your alarm clock, were late for class, received an F on your English paper, and got a flat tire, a trivial event such as stepping in some gum can evoke an intense emotional reaction.

There are two main biological components that play a role in perceiving emotion: the *autonomic nervous system* and the *limbic system*. When the sympathetic division of the autonomic nervous system is activated, the physiological changes that occur divert all of the body's energy toward escaping. When the parasympathetic nervous system is activated, a calming effect takes over the body. These autonomic responses are an integral part of the experience of an emotion. The **limbic system** is composed of a variety of structures located above the brain stem and below the cerebral cortex. Although many structures are involved in the functioning of the limbic system, there are four main structures responsible for emotion: the hypothalamus, the thalamus, the hippocampus, and the amygdala. The hypothalamus helps regulate the autonomic system by controlling the endocrine system and regulating the release of hormones. The thalamus is the primary relay station for almost all sensory input in the brain. The hippocampus is the structure in the brain responsible for memory formation, and specifically transferring short-term memories to long-term memory storage. The amygdala is responsible for assessing especially stimulating events as well as forming emotional memories. Stimulation leads to intense emotions, especially fear aggression, and violence, while destruction leads to tameness and a mellowed affect. **Kluver-Bucy syndrome** arises from the bilateral destruction of both amygdalae, resulting in a flat affect, hyperorality (compulsion to put things into one's mouth), hypersexuality, and behavioral disinhibition. The amygdala is also largely responsible for emotional learning.

The *prefrontal cortex* plays an important role in the cognitive component of emotion. This pathway is considered to be the slower of the two pathways, because it involves higher-order processing, unlike the instantaneous response from the amygdala. In general, the prefrontal cortex functions in decision-making, problem solving, and behavioral control. Our prefrontal cortex controls how we interpret a situation and thus, how we emotionally respond to it. Imagine the amygdala as the devil sitting on one shoulder and the prefrontal cortex as the angel sitting on the other shoulder. Although one is telling you to follow your initial gut feelings, the other is telling you to think logically and act in a socially appropriate manner.

The six universal emotions are happiness, surprise, fear, sadness, anger, and disgust, with characteristic, universal facial expressions associated with all of them. Unlike emotions **moods** are of much longer duration and are typically of lesser intensity. Moods are also not directed at a specific object. Evidence of universal, basic emotions suggests that emotions serve an adaptive role in evolution—emotions likely developed progressively over time as a communicative function.

Stress is defined as a state of mental and physical tension caused by a threatening or demanding situation. The largest factor that determines whether a situation is stressful or not is our appraisal of it. Our **appraisal**, or evaluation, of life's events is based on experience, instinct, and personality. **Ambient stressors** describe global challenges that affect individuals on a subconscious level, such as pollution, crime, city noise, and cultural expectations. They are non-urgent, negative, and chronic. When a person is faced with a challenge or threat, the body transitions to either the *fight-or-flight response* or the *tend-and-befriend* response. The **General Adaption Syndrome** is a stress model to explain three distinct stages of the stress reaction. The first stage, known as the **alarm phase**, is the initial reaction to the presence of a stressor. The threatening stimulus has evoked the body's fight-or-flight response. The second stage is the **resistance phase**, in which the body is flooded with stress hormones. Although not as intense as the immediate arousal of the alarm phase, the body continues its response as the body gradually grows weaker. If the resistance phase is not followed by recovery, the body enters the third and final stage known as the **exhaustion phase**—in this stage, the body becomes drained of its energy store and loses its ability to respond to the stressor. It becomes susceptible to illness, tissue and muscle damage, and possible death.

The **hypothalamic-pituitary-adrenal axis**, also known as the **HPA axis**, controls the stress response and mediates the general adaptation syndrome. The HPA axis is a circulatory system of hormonal controls between the hypothalamus, pituitary glands, and adrenal glands. It is responsible for both the release of glucocorticoids to inhibit the immune system and presence of glucose to focus all attention on a stressor. The process begins in the paraventricular nucleus of the hypothalamus by releasing corticotropin-releasing hormone (CRH) in response to a stressor. CRH acts on the anterior pituitary gland and tells it to release adrenocorticotropic hormone (ACTH), which in turn acts on the adrenal cortex, stimulating the release of glucocorticoid hormones (specifically cortisol) into the bloodstream. The release of cortisol initiates a negative feedback loop, acting back on the hypothalamus and the anterior pituitary to suppress the release of CRH and ACTH. **Cortisol** is a major stress hormone that suppresses the immune response and breaks down glycogen into glucose for energy, contributing to the overall sympathetic nervous system response.

Since cortisol increases glucose levels in the blood, prolonged exposure to a physiological stressor is associated with too much available energy, and chronically high blood glucose levels can result in **diabetes**. Cortisol also weakens the immune system by preventing inflammation, making the body more susceptible to infection. The heart also has to pump with greater force in order to increase the exchange of oxygen and nutrients—maintaining this high blood pressure can eventually lead to **hypertension** and other vascular diseases. Stressors also decrease the secretion of gonadotropin-releasing hormone (GnRH), which is responsible for stimulating the release of FSH and LH in females. This results in lower levels of estrogen and progesterone, as well as irregular menstrual cycles.

Psychological stressors lack a discrete onset and termination, and they originate in the mind. Chronic stress is associated with an increased risk of developing mental health problems, including depression, anxiety, OCD, phobias, and PTSD. **Coping** and **coping mechanisms** refer to the various conscious strategies aimed at reducing or managing stress levels. There are three types of coping strategies:

1. **Appraisal-focused strategies** are directed at modifying one's perception of stress. How a person thinks, and what they expect from themselves and others, influences the amount of stress they experience on a daily basis.
2. **Problem-focused strategies** are directed at modifying the stressor itself. They attempt to find the source of the problem and deal with it by learning new skills.
3. **Emotion-focused strategies** are directed at reducing the emotional response to a stressor. The negative emotions accompanying stress are largely what make stress such an adverse experience.

Techniques that relax both the body and mind can also be used to manage stress. Physical activity releases **endorphins**—natural painkillers—into the body that elevate mood, decrease tension, and improve sleep. Exercise also reduces the levels of cortisol in the body and increases cardiovascular health. Meditation can also manage stress.

Motivation is the process that gives behavior its energy and direction. An **instinct** is an innate behavior we do without forethought that controls each of our wants and desires. The **drive-reduction theory** is the idea that a physiological need creates a state of tension or arousal that drives the organism to reduce the need through performance:

$$\text{Need} \rightarrow \text{Drive} \rightarrow \text{Performance} \rightarrow \text{Goal} \quad (86)$$

A **need** is an excess or deficiency related to homeostatic balance or survival (e.g. food deprivation). The **drive** is the intervening variable that energizes behavior (e.g. hunger). The goal is some commodity (e.g. food) or change (e.g. pain reduction) that reduces the drive. The basis of this theory is homeostasis. Furthermore, a need produces a **drive stimulus**, which becomes associated with the response that reduces that drive. For example, a hungry child is told there is a piece of

candy hidden somewhere on a book case. The child looks under every book until she finds the piece of candy in 210 seconds. She is then sent out of the room, and a new piece of candy is hidden under the same book. In her next attempt, she finds the piece of candy in only 86 seconds. By the ninth search, she finds the piece of candy in 2 seconds. Since the girl had a drive stimulus (hunger), she searched for the piece of candy in order to reduce this drive. The response of looking under the particular book was rewarded and thus the behavior was reinforced. By incorporating this principle of **reinforcement** into the driver-reduction theory, both the *energy* and also *direction* of behavior is accounted for.

The **incentive theory of motivation** proposes that a *pull* of external stimuli, rather than just the *push* of internal drives, can also motivate behavior. An **incentive** is the magnitude and anticipation of a reward. The **arousal theory of motivation** suggests that people are driven to perform certain actions in order to maintain an optimum level of arousal. What is optimum varies from person to person. The **Yerkes-Dodson Law** states that performance increases with increased arousal, but only to a certain extent. For example, if you have increased arousal the morning of the MCAT, you will probably score higher than you did on your practice exams due to increased alertness and concentration. However, if you become overly stressed and excessively aroused, you will have passed the certain extent that is beneficial. Your inability to focus and high level of anxiety will make it more difficult to score well.

The **humanistic theory of motivation** focuses on the individual and the role of a **self-concept**—how we perceive ourselves. This theory suggests that people are motivated by a desire to reach their full potential; they can do so only when their most basic needs are met. **Cognitive theories of motivation** suggest that thinking is sufficient to explain behavior. With a purpose clearly in mind, people have the ability to select one course of action over another. Therefore, cognitive theorists view motivation as a thought process, rather than an automated response. There are three main cognitive theories of motivation:

1. The **expectancy-value theory** proposes that a person is motivated both by their expectancy and value of a goal.
2. The **goal-setting theory** suggests that a person's drive to reach an end state, as well as the characteristics of a particular goal, motivate a performance.
3. The **attribution theory** proposes that motivation stems from peoples' perceptions of their own successes and failures.

People tend to attribute their successes to internal factors and their failures to external factors.

Pavlovian-Instrumental Transfer (PIT) is a process by which environmental stimuli can motivate instrumental behavior. There are two types of PIT: *specific PIT* and *general PIT*. Specific PIT happens when the presence of a conditioned stimulus (reward-paired) motivates a behavior to obtain that specific reward. General PIT occurs when a conditioned stimulus activates a general psychological system, which in turn motivates a behavior to obtain a different reward. For example, let's say we see a McDonald's wrapper on the sidewalk, thus activating a mental representation of a cheeseburger. Since there is a McDonald's on every street corner, we walk down the road and purchase a burger as a result of that activation. This is an example of a specific PIT. Now, if you see a McDonald's wrapper on the sidewalk but have class in ten minutes, you will behave according to general PIT. The wrapper activates a representation of a cheeseburger, which will activate a general psychological system that motivates you to purchase a bag of chips on campus. Since a bag of chips is not associated with a McDonald's wrapper but is associated with satisfying your hunger, your response was general, not specific.

Biologically, the **nucleus accumbens** is the brain's reward center; it is found in the ventral striatum and experiences an increase in dopamine levels with food rewards, social interactions, sexual stimuli, and abused drugs. When dopamine agonists are injected into the nucleus accumbens, the motivational value of a stimulus is increased; in other words, the particular reward becomes more pleasurable. Therefore, dopaminergic input to the nucleus accumbens mediates the ability of reward-paired cues to motivate behavior.

To recap, motivation is the process that gives behavior its *energy* and *direction*.

Addiction is a chronic, often relapsing brain disease that causes compulsive drug-seeking and use despite harmful consequences to the addicted individual and those around them. There are four main theories of addiction:

1. The **allostasis theory**, also known as the **physical-dependence theory**, is the idea that addiction results from a process of maintaining apparent homeostatic stability through a change, but at a price. When a person takes a drug, it produces a positive emotional experience that is excessively beyond the set point for one's mood. In order to maintain homeostasis, the body triggers a correction response called the **opponent-process**. This adaptation opposes the initial emotional effects of the drug by slowly developing a counteractive process that gradually becomes larger, and eventually, completely masks the initial effects that the drug once had. Over time, the hedonic set point of the drug-user starts to decrease—think of your body as overcorrecting a problem. Drug addicts start to take the drug just to reach a feeling of being normal, rather than to feel the strong euphoric effects. The allostasis theory states drug use is driven by negative reinforcement—it is less about getting high and more about not feeling bad.

2. The **associative learning theory** of addiction is based on the idea that cues in the environment can come to elicit addictive behavior. The principles of classical conditioning suggest settings and drug-related stimuli can compel an addict to seek drugs, causing them eventually to relapse, despite the negative consequences.
3. The **incentive sensitization theory** of addiction states that exposure to psychoactive drugs changes the brain circuitry that is normally responsible for the pleasurable effects of incentives. The pleasurable incentive value of drugs increases with repeated exposure in certain individuals.
4. The **fronto-striatal dysfunction theory**, also known as **impulsivity**, suggests that drug-induced prefrontal cortex dysfunction leads to an inability to inhibit maladaptive behavior. The *striatum* is the region of the brain responsible for coordinating bodily movements with motivation, while the *prefrontal cortex* is the region responsible for planning, decision-making, and behavioral control. When drugs are abused, the frontal cortical regions responsible for inhibiting inappropriate behavior are damaged. In other words, exposure to drugs results in increased impulsive behavior that urges people to engage in risky events, despite knowledge of their negative consequences.

Psychoactive drugs produce qualitative changes in the conscious experience. There are three types of consciousness-altering drugs:

1. **Depressants** slow down CNS activity by increasing the activity of GABA, the main inhibitory neurotransmitter in the brain, and by decreasing the activity of glutamate, the main excitatory neurotransmitter in the brain. They typically have a calming, relaxing effect on behavior. Examples are alcohol, sedatives, and opioids (which are typically used for their analgesic effects of pain relief).
2. **Stimulants** increase CNS activity by enhancing the effects of norepinephrine, serotonin, and dopamine in the brain through various mechanisms. Almost all stimulants work by releasing dopamine and serotonin or by blocking their re-uptake within the synapse. Nicotine is a very common example, and acts on the CNS by binding to nicotinic cholinergic receptors in the brain and stimulating the production of epinephrine and dopamine. Caffeine, cocaine, amphetamines (adderall), and MDMA (ecstasy) are also examples of stimulants.
3. **Hallucinogens** cause distortions in one's perceptions of reality. LSD is a type of hallucinogen that results in a distorted emotional experience based on the dosage, the user's personality, and the surrounding at the time of ingestion. THC, the main psychotropic ingredient in marijuana, is classified as a mild hallucinogen. It affects the nervous system by over-activating the endocannabinoid system.

6.5 Cognition and Consciousness

According to the **information processing model**, the *sensory memory/sensory register* records sensory input for less than a second. From the sensory memory, information is processed in **working memory** (analogous to the CPU of the computer) it is eventually stored in long-term memory.

Attention is the process of concentrating on a select amount of information from a wide range of available information. Attention can refer to sensory input, memories, and thought processes. There are two types of attention: *selective attention* and *divided attention*. In **selective attention**, a choice is made to concentrate on a particular stimulus and to ignore other, less salient stimuli. The most common example of selective attention is described in Colin Cherry's **cocktail party effect**. When you are attending a cocktail party, there are numerous conversations occurring throughout the room, yet you are somehow able to focus on your own. The ability to tune out surrounding conversations shields the mind from potential distractors and allows us to allocate our mental resources efficiently.

Another example of selective attention is **searching**, or scanning the environment for a target stimulus. The more **distractors** (irrelevant stimuli that divert attention) that are present, the more difficult the search becomes. There are four theories to describe how easily a target stimulus can be found: *display-size effect*, *feature-integration theory*, *similarity theory*, and *guided search theory*. The **display-size effect** refers to the direct relation between the number of distractors in an array and the time required to locate a target stimulus among them. The **feature-integration theory** states that it is easier to perform a feature search than it is to perform a conjunction search. In a *feature search*, the target stimulus is found by scanning for a single, distinct feature among the distractors. In a *conjunction search*, the target stimulus is found by scanning for a combination of features among the distractors. The **similarity theory** of search states that the more similar the target stimulus is the distractor stimuli, the more difficult the search will be, regardless of the number of features. Lastly, the **guided search theory** proposes that all searches require two stages: developing a mental representation of the

target and evaluating all of the activated elements to find the target. These are all pretty intuitive when you're thinking about tasks such as finding Waldo and stuff like that.

In **divided attention**, your concentration is split between performing two or more tasks or inputs. Whether divided attention is easy or hard to perform depends upon the resources required. If the combined demand for resources is greater than the amount of resources available, then the two tasks will interfere with each other.

Inattentional blindness, also known as **perceptual blindness**, refers to the inability to see something in plain sight because of attending to another stimulus. A similar phenomenon is called **change blindness** and refers to a person's inability to detect visual changes in a scene he or she is directly looking at. Quite separately, the **Stroop effect** is a phenomenon that demonstrates a problem with selective attention. Once a controlled task becomes automatized through practice, there are not only benefits but also errors associated with this increased processing speed. For example, in a study done by Stroop (1935), participants were asked to state the ink color of each word in a list, out loud, as quickly as possible. The task is rather tricky; while you are attempting to focus on the ink color of each printed word, the meaning of the word is interfering with the task.

Attention deficit hyperactivity disorder (ADHD) is a psychological condition characterized by a failure to give close attention to details as well as an inability to sustain attention. The three main symptoms of ADHD are hyperactivity, impulsivity, and inattention. **Spatial neglect** is another type of attentional dysfunction characterized by a lack of attention paid toward a particular visual field. This dysfunction is often associated with a lesion located on the contralateral aspect of the brain (i.e. a lesion on the left hemisphere is associated with an inability to process information coming from the right visual field). **Fig. 6.5** displays drawing by a person with spatial neglect.

A **concept** is a mental representation through which we understand the world. Multiple concepts that have similar characteristics are grouped together and organized into a **category** within the mind. **Natural categories** are groupings that occur in the natural world and have predictable properties (e.g. trees, clouds), whereas **artifact categories** are human-made grouping with ambiguous, imprecise boundaries (e.g. furniture, clothes). How a concept becomes part of category is explained by several theories:

1. In the **feature-based theory**, concepts are placed into categories based on their defining features. For example, in order to be included in the category *husband*, you must possess the features of male and married.
2. In the **prototype theory**, concepts are grouped together by their degree of similarity to a **prototype**, which is the ideal or average model based on what we have previously encountered. Rather than specifying the boundaries, a concept is considered in the theory to part of a category based on its **graded membership**—objects that are more similar to the prototype are better members.
3. In the **exemplar-based reasoning theory**, concepts are grouped together by their degree of similarity to an **exemplar**, which is a specific remembered instance that is most typically thought of when a category is mentioned.
4. In the **theory-based view of meaning**, a concept is placed into a category based on a general idea constructed from experience. The theory-based view of membership depends on an explanation rather than a word or two. Since this theory does not base categorization on resemblance, it accounts for the reason why a counterfeit bill is not considered money, despite having a nearly perfect resemblance to real money.

Knowledge in the form of concepts and categories can also be organized within the mind as **schemas**, which are broad patterns of what is normal in a given situation. A **stereotype** is a fixed, generalized belief about a particular concept. A **script** is an example of a schema that contains information regarding the order in which things should occur. For example, most people have a script for a restaurant that tells them how to behave and what to expect.

The **problem-solving cycle** is the series of steps that lead us to a solution for a problem. It involves identifying the problem, defining the problem, forming a strategy, organizing information, allocating resources, monitoring progress, and evaluating results. The following is a list of various approaches to problem-solving:

1. **Algorithms** are sequences of steps that always produce a correct answer. Although effective in solving a problem, algorithms are not necessarily always the 'best' strategy—they can be both time- and resource-consuming.
2. **Heuristics** are mental shortcuts that do not guarantee a correct answer but effectively narrow the problem space (the set of all possible states that can be reached to obtain a solution) and simplify the problem in order to speed up the process. There are two types of heuristic problem-solving strategies. In the **hill-climb strategy**, you simply choose the option that moves you directly towards your goal. The drawback of this approach is that sometimes you need to move backwards to go forward. In **means-end analysis**, you compare the current state to the goal state and then try to decrease this distance with the available means.

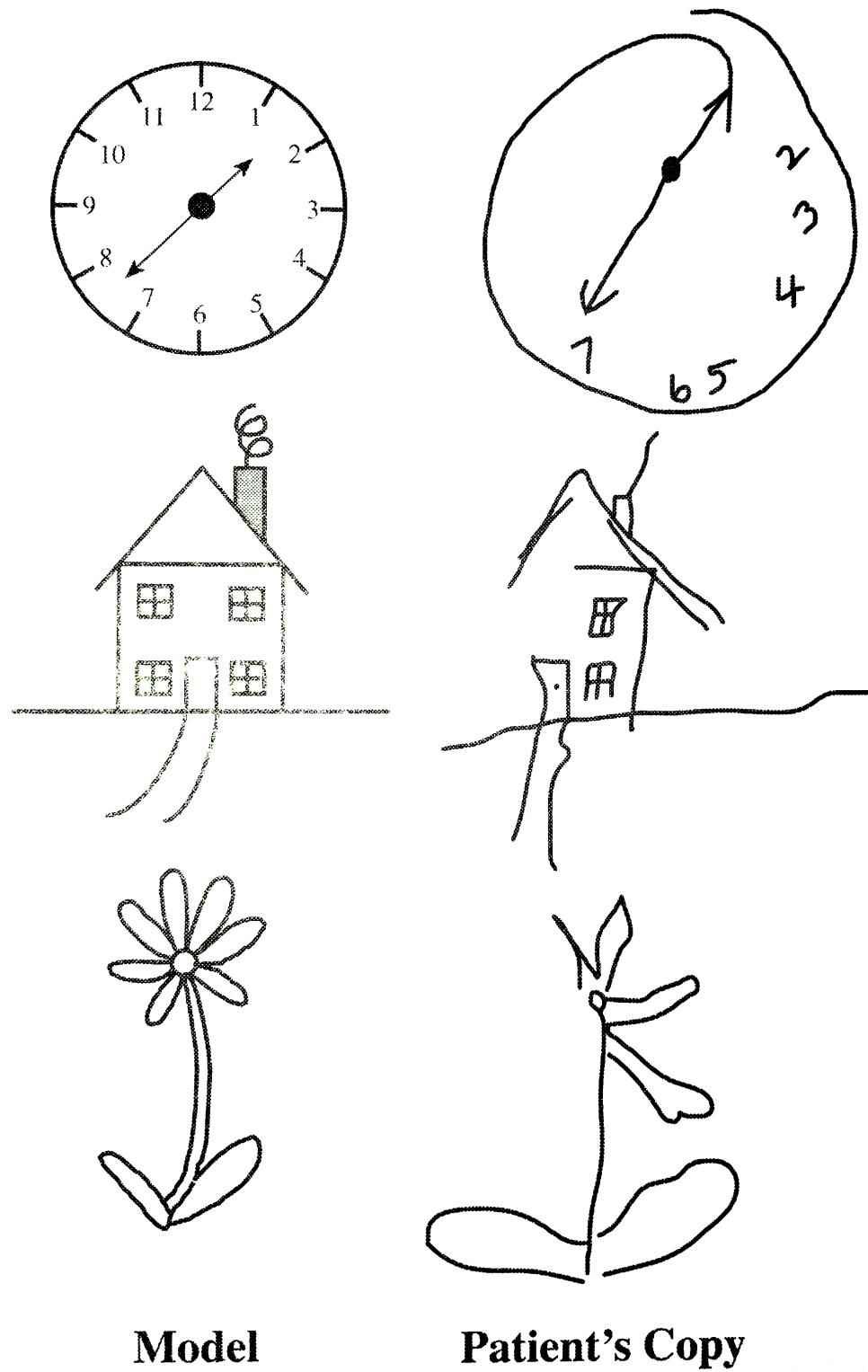


Figure 113: Spatial neglect. The lesion is on the right hemisphere in this case.

3. In the **analogy** approach to problem-solving, a current problem is compared to a similar, already-solved problem, and the solution is translated. With this strategy, it is the dynamic (deeper) traits of the solution that translate rather than the superficial (surface) features.
4. **Trial-and-error** is a problem-solving strategy in which all possible solutions are tried until one works. This technique is usually beneficial only when the problem space is relatively small.
5. When applied strategies fail, sometimes **insight**—a sudden, novel solution to a problem—succeeds. Insight is often described as the ‘aha moment’ in psychology.

A **problem-solving set**, also known as a **mental set** or **Einstellung**, is the collection of beliefs and assumptions that a person makes about a problem based on previous experiences. This is a common reason to explain why a problem cannot be solved. Another influence is the tendency to be rigid in how we think about an object’s function, known as **functional fixedness**. Owing to this barrier to problem-solving, people’s past experience with an object can prevent them from using the object in novel ways. Additionally, a type of cognitive tendency in which people search for information that confirms their preexisting beliefs is **confirmation bias**. In other words, people tend to seek out solutions that are simple and agree with what they already think is probably true. **Belief perseverance** is the tendency to stick with one’s initial beliefs, even in the face of new and contradicting evidence. Other barriers to problem-solving include making assumptions and getting sidetracked by irrelevant information.

Convergent thinking narrows down the possibilities to find one answer, whereas *divergent thinking* attempts to generate a variety of ideas to a problem. People who employ divergent thinking tend to come up with more creative ideas. **Wallas’ stages of creativity** attempts to explain the creative approach to solving problems:

1. In the **preparation stage**, information is gathered about the problem. This is the collection of mental resources that will be used to plan, organize, and guide the problem-solving process.
2. In the **incubation stage**, the problem is momentarily set aside and processed unconsciously. Although there is no direct effort or attention paid to the current problem, the brain’s mental processes are still at work.
3. In the **illumination stage**, a key insight into solving the problem is suddenly realized. It is the conscious culmination of the many cognitive processes at work in stages one and two.
4. In the **verification stage**, the creative thinker confirms that the new idea does lead to the problem’s solution.

The most basic model used to describe the cognitive processes of decision-making stems from an economic perspective. People calculate expected value = (subjective probability) \times (subjective utility) and want to maximize the expected value. Beyond risk-benefit analysis, decision-making is based on the mental shortcuts of heuristics and biases. As we previously indicated, **heuristics** are mental strategies that sacrifice a 100% guarantee of accuracy in exchange for efficiency. There are two main types of heuristics in decision-making:

1. With the **availability heuristic**, a judgment is made based on the perceived frequency or likelihood of an event. For example, people often think car accidents cause more death in America than stomach cancer (even though stomach cancers kill many more people) because car accidents are more often mentioned in the mass media and are therefore more ‘available’ in the mind. The **conjunction fallacy** is the mistaken belief that a smaller, specific subset of a category is more probable than the larger, general set.
2. With the **representative heuristic**, one’s judgment of an event is based on its similarity to the prototype of the population and its perceived randomness. In other words, the more closely an event appears to resemble the standard and the more random it appears, the more likely it is to be considered probable. An example of this can be demonstrated with the **gambler’s fallacy**, where people mistakenly believe that the previous outcomes at a casino will affect the likelihood of their own winning or losing.

Cognitive biases describe a tendency to think and perceive something in a particular way. Like heuristics, biases function to decrease the effort required to make a decision at the expense of errors. There are three biases that occur very frequently:

1. **Illusory correlation** is the phenomenon by which people perceive a relationship between two unrelated variables.
2. **Overconfidence** is a bias in which a person’s evaluation of his or her own judgments is greater than the actual accuracy of those judgments. For example, researchers have found that when people say they’re 100% confident, they are right only about 80% of the time.

3. **Hindsight bias** is the idea that, when one evaluates a past event, he or she feels that the results were always predictable and should have been obvious. This is also known as the ‘knew-it-all-along effect.’
4. Somewhat related, the **framing effect** is a change in how a problem is presented that affects decision-making. Generally speaking, when making a choice, people tend to seek risks with negative framing and avoid risks with positive framing.

There are three cognitive theories of intelligence:

1. The theory of multiple intelligences describes eight different intelligences that function independently of one another: linguistic, logical-mathematical, spatial, musical, bodily-kinesthetic, interpersonal, intrapersonal, and naturalist. **Savant syndrome** is a condition in which a person is exceptionally advanced in one area of functioning but suffers from severe cognitive deficits.
2. The triarchic theory of intelligence divides intelligence into analytical abilities, creative abilities, and practical abilities.
3. The three-stratum model of intelligence arranges intellectual ability in a hierarchy consisting of three strata. The first, bottom layer is called the narrow stratum, which includes specific abilities such as reading comprehension and spelling ability. The second layer is the broad stratum including short-term memory, retrieval ability, and information processing speed. The third, top layer is called the general stratum, which contains the *g factor*, which is measured performance on a verbal reasoning task or a spatial, mathematical task.

The IQ is the **intelligence quotient** of a person, and is defined as the percentage ratio of a person’s mental age divided by chronological, physical age.

There are four stages of cognitive development. Here they are in chronological order:

1. The **sensorimotor stage** typically ranges from birth to the acquisition of language, which is typically around two years old. During this stage, babies attempt to understand the world through exploring it; they touch, suck on, listen to, and look at everything around them. The main developmental phenomenon that appears in this stage is **object permanence**—the awareness that objects continue to exist even when they are not perceived. At around 8 months old, babies also start to develop **stranger anxiety**. As they start to become more socialized, babies often show fear, anxiety, and distress toward unfamiliar faces.
2. The **preoperational stage** ranges from about two years old to six or seven years old. In this stage, children are not yet able to perform mental operations, such as reasoning, but they begin to think in terms of images. Children at this age are considered **egocentric**—they have difficulty perceiving others’ viewpoints. Another developmental phenomenon that characterized this stage is an inability to understand the concept of **conservation**—the quantity of a substance remains unchanged despite changes in its shape. Near the end of this stage, children begin to develop a **theory of mind**—an ability to infer another person’s mental state and thus explain and predict their behavior.
3. The **concrete operational stage** lasts from about seven years old to eleven years old. This stage is characterized by the ability to think logically about concrete events but the *inability* to comprehend hypothetical or abstract concepts.
4. The **formal operational stage** begins around the age of 12 and proceeds through adulthood. At this stage, children are now able to think logically about abstract and hypothetical instances; they are capable of deductive reasoning.

In late adulthood, changes in cognitive development are multidirectional. Attention and memory generally tend to decline with age. One’s **crystallized intelligence**—information acquired over a lifetime through experience—continues to grow steadily. However, one’s **fluid intelligence**—the ability to apply basic information-processing skills—peaks in early adulthood and then declines. The culture, heredity/genetics, and environment also affect cognitive development throughout one’s life.

Biologically, people’s large cerebral cortex and numerous folds to increase surface area play a large role in cognition. **Neoplasms** (brain tumors), physical trauma/concussions, and **strokes** can cause severe cognitive dysfunction.

There are three major theories on language development that you should know about:

1. The **learning perspective** suggests that children develop the use of language through associative conditioning (classical and operant) and observational learning. Children listen to and observe the world around them.
2. The **nativist perspective** of language development stems from the idea that the entire complexity of language cannot be learned from experience alone; language is an innate skill for which people are predisposed. This perspective proposed the existence of a **language acquisition device (LAD)**, an area of the brain with a set of innate transformational grammatical rules. The rules produce different utterances for different languages, yet the potential to produce grammatical sentences lies in every human brain.

3. The **interactionist perspective** is based on the idea that children acquire new understandings that they wish to communicate with others. It proposes the existence of a **zone of proximal development (ZPD)**, or an area of optimal learning in which a child develops language skills through the child-adult interaction. The interactionist perspective recognizes the interplay of environmental and biological factors, while stressing the social support aspect of this interaction.

Language relativity is the idea that speakers of different languages utilize different cognitive processes that influence how they think. This concept is also known as the *Sapir-Whorf hypothesis*, and contends that language is a reflection of how a culture thinks about things. **Linguist determinism**, also known as the *strong Sapir-Whorf hypothesis*, is the idea that language entirely determines how we think. Although the linguistic relativity hypothesis is applicable in most instances, it is possible to think about concepts for which language does not have a word. For example, try to imagine a shade of the color blue for which there is no word. For most people, this task is possible.

As with most other cognitive abilities, language comprehension and the production of speech correspond to particular regions of the brain. Two areas of the brain that are strongly associated with language are *Broca's area* and *Wernicke's area*. Located in the left frontal lobe, **Broca's area** is responsible for speech production. Damage to this region of the brain results in the loss of the ability to speak words, but the ability to comprehend language is conserved. In other words, the person with this damage has developed a type of **aphasia**—a language disorder associated with brain injury. Located in the left temporal lobe, **Wernicke's area** is responsible for language comprehension. When this area of the brain is damaged, people maintain the ability to speak, but fail to understand the meaning of words. Their sentences usually don't make sense even though the individual words do.

Consciousness is the awareness that we have of our environment as well as our mental processes. Consciousness allows us to contemplate many things, such as reflecting on the past, considering consequences, planning for the future, or trying to grasp obscure concepts like infinity. There are four naturally occurring states of consciousness, which are summarized in **Fig. 6.5**:

1. **Alertness** is the state of consciousness characterized by being awake and is most likely the state we think of when considering consciousness. When one is alert, he is fully aware of their environment and their thoughts. On an EEG, the state of alertness is characterized by β waves—brain waves with a frequency between 13 and 30 Hz.
2. **Daydreaming** is the state of consciousness wherein one is awake, but not fully alert. Compared to alertness, the state of daydreaming leaves one more relaxed and less focused. Daydreaming is characterized by α waves—brain waves with a frequency between 8 and 13 Hz. Daydreaming can also be induced by meditation, a topic we will cover later in this section.
3. **Drowsiness** is the state of consciousness with even less awareness, and often experience either just before falling asleep or just after waking up. This transition period is characterized by θ waves—brain waves with a frequency between 3 and 8 Hz.
4. **Sleep** involves the inhibition of most voluntary muscles and sensory activity. Deep sleep is characterized by δ waves—brain waves with a frequency between 1 and 3 Hz.

Sleep has five distinct stages. Marking the onset of sleep, **stage 1** lasts roughly 20 minutes and is characterized by decreased α waves and increased θ waves. During this period, we are most easily awakened by and often experience *hypnic jerks*—involuntary muscle twitches associated with the sensation of falling. **Stage 2** involves the decrease of body temperature and slowing of breathing and heart rate. Additionally, stage 2 is characterized by increased θ -waves and two distinct waveforms: *K-complexes* and *sleep spindles*. *K-complexes* are single, large-amplitude, high-voltage waves that are believed to suppress cortical arousal and aid sleep-based memory consolidation; *sleep spindles* are short bursts of 12-14 Hz waves that also suppress cortical arousal. From here, one falls into the deeper, *slow-wave sleep* of stages 3 and 4. **Stage 3** is characterized by decreased θ waves and increased δ waves, and **stage 4** is almost entirely δ waves. Generally, there is a gradual decrease in brain wave frequency and an increase in brain wave amplitude, as well as an increase in brain wave voltage, as one progresses from stage 1 to stage 4.

After remaining for a time in stage 4, the sleeper retreats in reverse order through the other stages of sleep back to (but not including) stage 1. But instead of waking up, the sleeper enters **REM (rapid eye movement) sleep**, a stage of sleep characterized by low-amplitude, mixed-frequency brain waves very similar to those of an alert and actively thinking person. The average adult typically experiences four sleep cycles throughout the night. And as the night progresses, the REM sleep stage gets longer and longer, while the non-REM stages get shorter and shorter (particularly stage 4). Dreaming

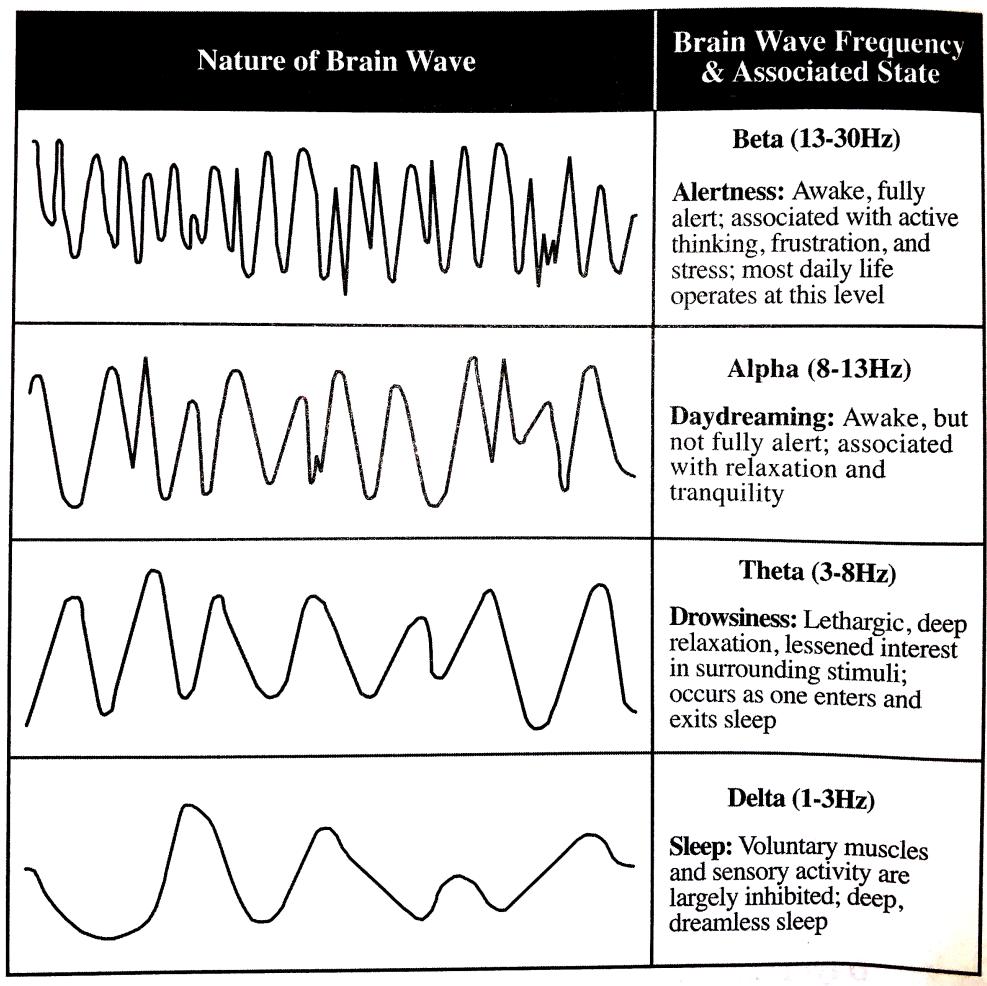


Figure 114: The four major types of brain waves and their associated states of consciousness.

occurs predominantly in REM sleep. Some believe that dreaming and REM sleep provide regular brain stimulation that helps develop and preserve neural pathways. For some reason, studies have shown that REM sleep is physiologically required. There are three main, complementary types of theories of sleep:

1. *Adaptation theories* contend that human beings evolved to sleep during the night time to conserve energy and to avoid predation.
2. *Recuperation theories* argue that being awake disrupts the body's homeostasis and that sleep is necessary to restore it.
3. *Consolidation theories* propose that sleep is essential to maintain proper *brain plasticity*—changes in the structure and organization of the brain.

Sleep, along with many other biological processes, is regulated according to internal biological timing mechanisms called **circadian rhythms**, controlled primarily by light. Located in the retina are specialized ganglion cells that respond directly to light and project to a small cluster of neurons in the hypothalamus called the **suprachiasmatic nucleus (SCN)**, the circadian control center. The SCN processes information concerning the length of day and night and signals to the **pineal gland**, which is dorsal to the thalamus, either to increase or decrease its production of the hormone **melatonin**. Increased levels of melatonin cause the body to become more relaxed and sleepy. The absence of light information from retinal ganglion cells ultimately signals to the pineal gland to increase melatonin release; the presence of light signals to inhibit melatonin release.

Sleep deprivation often results in irritability and in impaired learning and memory function, as well as many other

unfavorable effects on behavior. There's **insomnia** where you struggle to fall or to remain asleep, and also **hypersomnia** where you sleep excessively; insomnia is much more common. **Sleep apnea** is a condition where the person periodically stops breathing and suddenly awakens to gasp for air before falling back asleep. This cycle can happen around 100 times per night. Most patients with sleep apnea are unaware of this cycle and instead of complain of increased sleepiness. Meditation, hypnosis, and conscious-altering drugs can alter the natural states of consciousness.

6.6 Personality and Attitudes

Personality encompasses one's characteristic patterns of feelings, thoughts, motivations, and behaviors that make him or her unique. Personality typically remains fairly consistent throughout life, predictably guiding one's actions in different situations and over extended periods of time. There are six major theories of personality and their approach to explaining the development of personality and its effect on behavior. These theories include psychoanalytic, humanistic, trait, social-cognitive, behaviorist, and biological theories.

Psychoanalytic theories emphasize the dynamic interaction between the conscious and unconscious mind on the development of personality and the discourse of behavior. Specifically, Sigmund Freud believed that human personality arises from **conflict**—the struggle between our pleasure-seeking, aggressive, biological impulses and our internalized social restraints. To explain the dynamic interactions behind this conflict, Freud divided personality into three interacting components:

1. The **id** is unorganized and operates entirely unconsciously, constantly seeking to relieve tension associated with hunger, sex, aggression, and other primal urges. The id functions according to the *pleasure principle*—it seeks immediate gratification regardless of external factors or consequences.
2. The **ego** operates both unconsciously and consciously to compromise realistically with the id and the outside world. The ego acts according to the *reality principle*—delaying immediate gratification in order to weigh its costs and benefits appropriately.
3. The **superego** is predominantly unconscious, and instills a sense of conscience (moral compass) that forces the id's desires to be met not only realistically, but also morally. The superego judges one's actions and is responsible for positive feelings (e.g. pride) or negative feelings (e.g. guilt). Oftentimes the id's impulsive demands oppose the superego's restraining ones, causing the ego to act as an executive mediator between the two.

Freud argued that the id, ego, and superego are continually in conflict with one another, and that if not effectively mediated by the ego, this conflict generates anxiety. Furthermore, if this anxiety persists, individuals become too overwhelmed to complete everyday tasks. Freud postulated that when the ego feels itself succumbing to this internal struggle, the ego protects itself via **defense mechanisms**—tactics that redirect or reduce conflict and anxiety by distorting reality. All defense mechanisms function on an unconscious level, much like our immune system.

Freud also postulated that personality develops during the first few years of life as children pass through a series of **psychosexual stages**—stages in which the pleasure-seeking impulses of the id become focused onto pleasure-sensitive body parts called *erogenous zones*. The psychosexual energy driving these impulses and behaviors is called **libido**. There are five psychosexual stages: oral, anal, phallic, latency, and genital. A 'normal' personality results if these psychosexual stages are completed successfully. If an conflict remains unresolved during a particular stage, **fixation**—the persistent focus of libido to that stage—can occur, sometimes surfacing as maladaptive behavior during adulthood. For example, an infant who is weaned prematurely from breastfeeding during the oral stage may become fixated at that stage and develop smoking or nail-biting tendencies as an adult. The five psychosexual stages are described here:

1. From birth to around 18 months of age, the **oral stage** characterizes a time where the infant's primary source of interaction occurs through the mouth. Accordingly, the infant is focused on oral pleasures associated with sucking and rooting (the instinctive behavior to find the nipple) as he or she develops trust and comfort. The major conflict during this stage is weaning. Too much or too little oral gratification from weaning at an improper time can result in an oral fixation that may manifest in issues with dependency or aggression. Oral fixations are also believed to predict behavioral problems associated with eating, drinking, smoking, or nail-biting.
2. The **anal stage** typically occurs between 18 months and 36 months of age and characterizes a time where the libido's primary focus is pleasure associated with bowel and bladder elimination. The major conflict during this stage is toilet training—successful completion develops control and independence associated with creativity, competence, and productivity. Freud contends that societal pressure, mainly from parents, can be either too lenient—resulting in an

anal-expulsive personality, in which the individual is disorganized or destructive—or too stringent—resulting in an *anal-retentive* personality, in which the individual is obsessive and controlling.

3. From 3 to 6 years of age, the **phallic stage** is characterized by a libidinal shift in focus to the genitals. Boys develop unconscious sexual desires for their mothers while simultaneously developing jealousy and hatred for their fathers—a phenomenon termed by Freud as the **Oedipus complex**. Other psychoanalysts believed that girls experienced a similar **Electra complex**. Eventually, the child begins to repress these threatening feelings while also identifying with the same-sex parent as part of a coping strategy. Psychoanalysts argue that this identification stage is crucial for a child's development of **gender identity**. A phallic fixation could result in either the overindulgence or avoidance of sex as well as a weak or confused gender identity.
4. From age 6 until puberty, the **latent stage** characterizes a period of ego and superego development as well as libido suppression. This is a phase where individuals focus on hobbies and peer interests while their sexual feelings typically lie dormant.
5. From puberty onward, the **genital stage** is accompanied by a maturation and intensification of sexual interests, generally aimed at members of the opposite sex. The primary libidinal focus during the final psychosexual stage is the genitals.

Humanistic theories focus on the different ways normal, ‘healthy’ people utilize free will to develop personality and strive for self-realization and self-determination. These theories starkly contrast with Freud’s study of ‘sick’ people and the predetermined motives acquired during childhood that would shape their personalities. Typically optimistic in nature, humanistic theories focus on one’s ability to think consciously and rationally while acting with self-control. According to this theory, a hierarchy of needs governs our motivations, and more specifically, we all strive to satisfy the highest need of **self-actualization**—fulfilling our potential and finding purpose. There are three essential environmental conditions required to promote one’s growth successfully: genuineness, acceptance, and empathy. **Genuineness** allows one to be transparent with his or her feelings; **acceptance** fosters an attitude of *unconditional positivity* that teaches self-worth despite individual shortcomings; **empathy** allows people to share and understand each other’s feelings. There’s also the idea of one’s **self-concept**, which is basically the image that you have of yourself. If there is incongruence between your self-concept and reality, then you could be insecure, dissatisfied, and unhappy.

Trait theories are pretty neutral compared to the negative approaches of psychoanalytic theories and positive approaches of humanistic theories. They describe personality in terms of **traits**—relatively stable characteristics that predictably guide behavior. They focus on describing individual traits, rather than explaining why they occur. There are four primary trait theories that we will focus on:

1. In **Allport’s Trait Theory**, there are three types of traits. **Cardinal traits** (e.g. altruism and ambition) dominate an individual’s life and shape his or her behavior. The American psychologist Gordon Allport suggested that these traits develop later in life and characterize the most evident and persistent aspects of personality. **Central traits** (e.g. honesty and kindness), while not as dominant as cardinal traits, are the general characteristics found in varying degrees from person to person. **Secondary traits** (e.g. stage fright and love of the outdoors) only present themselves under specific circumstances and refer to the detailed, less obvious aspects of one’s personality.
2. Trait theorist Raymond Cattell took the list of all traits and narrowed it down to 16 key personality traits, but eliminating uncommon traits and identifying clusters of test items using **factor analysis**. The resulting *Sixteen Personality Factor Questionnaire* is one of the most widely used personality assessments to date.
3. **Eysenck’s 3 Dimensions of Personality** focused on just three universal traits: *extraversion-introversion*, **emotional stability-instability (neuroticism)**, and *psychoticism*. These are pretty self-explanatory except for **psychoticism**, which refers to an individual’s difficulty in dealing with reality. A psychotic person might be considered manipulative, antisocial, nonempathetic, or hostile. Depending on the combination of one’s extraversion-introversion and emotional stability-instability, his or her personality is categorized as *melancholic* (unhappy and depressed), *choleric* (angry and hot-tempered), *phlegmatic* (dull and unemotional), or *sanguine* (passionate and cheerful).
4. The **Big Five-Factory Theory** postulates that there are five core traits that interact to produce human personality: openness, conscientiousness, extraversion, agreeableness, and neuroticism. When trying to remember these five factors, think of the acronym OCEAN, and then remember that the OCEAN is big.

Social-Cognitive Theories of personality emphasize the collaboration between an individual's traits and his or her current situation. The *social* aspect of these theories posits that humans learn many of their behaviors through traditional conditioning or observing and imitating those around them. The *cognitive* aspect of social-cognitive theories indicates that our thoughts and perceptions concerning a particular situation affect our behavior. There are also **reciprocal determinisms**, where behavior is influenced by and influences one's social environment and cognitive processes. Social-cognitive theories also emphasize the individual's sense of **personal control**—whether the individual believes that he or she controls or is controlled by the environment. People who believe that chance or external factors decides their fate are said to have an **external locus of control**. These people often feel helpless and oppressed more frequently. People who believe that they control their own fate are said to have an **internal locus of control**. These people are more independent, achieve more academically and professionally, and cope with stress better.

Behaviorist theories explain personality through learning and suggest that personality results from an individual interacting with his or her environment. Behaviorist theorists, however, do not believe that the internal thoughts or emotions affect personality and study only physically observable or quantifiable behaviors. It is postulated that people have consistent personalities and behavioral patterns because they develop certain *response tendencies* over time. Behavior is formed through learning processes (i.e. *operant conditioning*) in which responses either result in positive or negative consequences that reinforce behavior. Personality develops and changes over the course of one's lifetime to maximize behavior associated with rewards and minimize behavior associated with punishments. In this sense, behaviorist theories contend that personality is a dynamic psychological concept that is influenced by the environment. There is also *classical conditioning*, which occurs when a conditioned stimulus is paired with an unconditioned stimulus, causing an association to form between the two. **Biological theories** focus primarily on how and why personality develops through evolutionary, genetic, and neurological perspectives. Personalities develop in line with natural selection and are adapted and selected for depending on social and environmental influences. Temperaments are also considered to be predisposed genetically. Brain-activity scans differ significantly between extraverts and introverts. One of the biggest strengths of biological theories of personality is their quantifiable objectivity. The six different personality theories that we have just discussed are summarized in **Fig. 6.6**.

Attitudes are learned tendencies to evaluate things with some degree of favor or disfavor. Attitudes can be polarizing

| Personality Theory | Hypothesis | Key Components | Criticisms |
|--------------------|--|--|--|
| Psychoanalytic | Personality is the result of fixation at various psychosexual stages and conflict between the conscious and unconscious; defense mechanisms prevent onset of anxiety | Id, Ego, and Superego Defense mechanisms Fixation Psychosexual stages | Very controversial Some claims disproven Overall negative focus Subjective, hard to prove |
| Humanistic | People are inherently good and strive for self-realization | Self-realization Positive environment (openness, acceptance, empathy) Self-concept | Overly optimistic and naive Subjective, hard to prove |
| Trait | Personality is the combination of relatively stable characteristics called traits | Allport's Trait Theory Cattell's 16 Personality Factors Eysenck's 3 Dimensions Big Five-Factor Theory | Traits are too consistent No explanation of why traits develop |
| Social-Cognitive | Personality results from the interaction of social environment, cognitive processes, and current behavior | Reciprocal determinism Personal control (external vs. internal locus control) | Too much emphasis on one's situation and/or environment Learning is different between people and ages |
| Behaviorist | Personality results from the interaction between the individual and environment; cognitive processes excluded | Operant conditioning Classical conditioning | Ignores internal traits and processes Generalizes animal studies to humans |
| Biological | The development of personality is largely influenced by biological processes | Evolution Genetics Neurological components | Not comprehensive Uneven focus on certain traits |

Figure 115: Personality theories.

or unifying; they can be strong or weak. **Explicit attitudes** influence our beliefs and behaviors on a conscious level of which

we are fully aware. **Implicit attitudes** also influence our beliefs and behaviors, but on an unconscious level. here are also three components simultaneously associated with each attitude:

1. The **affective component** encompasses one's emotions and feelings about the attitude object. For example, "I am afraid of bees." Notice the keyword that indicates emotion: 'afraid.'
2. The **behavioral component** encompasses the way in which an attitude affects how one behaves. For example, "If a bee buzzes toward me, I will run away." Notice the key phrase that indicates behavior: 'I will run.'
3. The **cognitive component** encompasses one's thoughts and beliefs about the attitude object. For example, "I think bees are dangerous." Notice the keyword that indicates cognition: 'think.'

Generally speaking, attitudes are either positive or negative. Three main types of learning influence attitudes: observational learning, operant conditioning, and classical conditioning. At some point, we have all developed or changed an attitude as a result of persuasion. When being persuaded, there are three main characteristics of a message that influence our evaluation of the message and, in turn, our attitude concerning the message. **Target characteristics** encompass many different features that influence the individual's interest in the message, such as intelligence, alertness, self-esteem, mood, and hunger, to name a few. **Message characteristics** include the various features of the message itself, ranging from its logic and flow to its length and vocabulary. **Source characteristics** encompass features of who and where the message came from, such as the source's expertise, trustworthiness, locational context, and attractiveness.

The **elaboration likelihood model (ELM)** of persuasion describes how attitudes form and change based on these three characteristics of messages. According to the ELM, people cognitively process a message first by target characteristics and then by message and source characteristics. Furthermore, this processing occurs along one of the two possible routes:

1. The **peripheral route** occurs when the individual has little interest in the subject and a low motivation and/or ability to think systemically about the message. When evaluating a message via the peripheral route, people are more likely to rely on superficial cues, such as source attractiveness or number of arguments, rather than message content or credibility. The peripheral route leads to shallow processing of a message and an attitude that either is unchanged or easily susceptible to further persuasion.
2. The **central route** occurs when the individual not only has a high interest in and motivation for the subject, but also the cognitive ability to evaluate the message critically. People evaluating a message via the central route are more likely to rely on the quality of the information presented.

Social psychologists have found that **attitude strength**—the degree to which an attitude is held—is often a good predictor of behavior. Stronger attitudes are more likely to affect behavior. We tend to assume that one's behavior is consistently aligned with his or her attitudes. In other words, we anticipate a predictable and rational link between attitudes and behaviors. The **theory of reasoned action (TRA)** posits that an individual's behavior is most accurately determined by his *intention*, or cognitive readiness, to perform the behavior. Furthermore, TRA suggests that this intention is a function of the individual's *attitudes* and perception of the *subjective norms* relates to the behavior, which in turn increases the likelihood of performing a particular behavior.

We have seen that people typically behave according to their attitudes. Perhaps more surprising is that the reverse is also true: behavior can affect attitudes. There are three related behavioral processes:

1. The **foot-in-the-door phenomenon** describes the tendency for people to be more likely to comply with a large or difficult request, if they have first agreed to a smaller or easier request. The key is to start small and build bigger.
2. **Role-playing** refers to the expected actions of the people who fill a certain social position. Research shows that people tend to internalize the roles that they play, modifying their attitudes throughout to fit the roles better. This was exemplified by the prison study experiment in 1972.
3. **Cognitive dissonance** is the psychological distress we experience by having conflicting thoughts or beliefs at the same time. The **cognitive dissonance theory** posits that we strive to reduce this tension, often by revising our thoughts and attitudes to make them more consistent with one another. Cognitive dissonance theory is especially pertinent for explaining discrepancies between behavior and attitudes. There are four main strategies that people utilize to eliminate or reduce cognitive dissonance: (1) acquire new information, cognitions, or attitudes, (2) reduce the importance of conflicting cognition or attitude, (3) modify or deny conflicting cognition or attitude, or (4) change the conflicting cognition, attitude, or behavior.

6.7 Psychological Disorders

A **psychological disorder** is the characteristic pattern or abnormal thoughts, feelings, or behaviors, which causes distress for the afflicted individual and impairs his or her daily life. There are two main types of approaches that affect how a particular disorder is treated: the *biomedical approach* and the *biopsychosocial approach*. The **biomedical approach** is the standard model used today for treating disease and assumes that all distress and disability are grounded primarily in certain biomedical factors. Thus, professionals who practice this approach focus on identifying and correcting the pathology to provide symptom relief. The **biopsychosocial approach** assumes that there are important psychological and social components to any disorder in addition to biological influences. This approach is a much more holistic and individualized approach used to diagnose and treat psychological disorders. Here are the most heavily tested disorders on the MCAT:

1. **Anxiety disorders** are related to anxiety, which is a state of inner turmoil characterized by nervousness, worry, and unease. There's *generalized anxiety disorder (GAD)* is a mental illness characterized by feelings of anxiety that are out of proportion to the individual's everyday problems. *Panic disorder* is marked by repeated panic attacks that cripple normal living; a panic attack is the unexpected, sudden onset of terror, an increased heart rate, faintness, weakness, and tingling in the hands. *Phobic disorders* are a type of anxiety disorder characterized by an extreme, irrational fear toward a specific stimulus. There's also **OCD** and **PTSD** which you're already familiar with. Generally, anxiety disorders are treated with a combination of medications and psychotherapy. These disorders are often accompanied by other disorders, such as substance abuse.
2. **Depressive disorders** are related to **anhedonia**, which is a loss of the capacity to experience pleasure. Although symptoms vary in frequency, severity, and duration, the most common symptoms include persistent sadness, hopelessness, irritability, loss of interest in activities that were once enjoyed, fatigue, changes in eating habits, decreased libido, and thoughts of suicide or suicide attempts. **Major/clinical depression** is a mental illness characterized by multiple severe symptoms from the list above that last for the majority of the time for at least two weeks. People suffering from major depression have a low mood and lack interest in activities that once seemed enjoyable. **Persistent depressive disorder**, also known as **dysthymia**, is a depressed mood that lasts the majority of the time for two or more years. **Bipolar disorder** is a condition characterized by cycling intense mood changes that often begin during late adolescence or early adulthood. A person suffering from bipolar disorder has moods ranging from extreme highs to extreme lows. Bipolar disorder is a lifelong mental illness that tends to recur intermittently. Of note, **cyclothymia** is a mild form of bipolar disorder characterized by episodes of hypomania and depression lasting for at least two years. There are two important biological theories of depression. The **monoamine theory of depression** contends that the mood disorder is associated with decreased activity of serotonin, noradrenaline, and dopamine. The fact that antidepressants are almost all agonists of these neurotransmitters supports the idea that depressive disorders are caused by deficits in monoamines. The **neuroplasticity theory of depression** proposes that the processes associated with a neuron's ability to adapt are disrupted in depressive disorders.
3. A **somatoform disorder** is a psychological disorder characterized by physical symptoms that cannot be attributed to a physical cause, substance abuse, or another mental illness. **Somatization disorder** is a long-term condition characterized by physical symptoms in more than one part of the body, without physical cause. **Conversion disorder** is a condition in which neurological symptoms appear without medical cause. It is believed to be the physical manifestation of psychological stress. **Hypochondriasis** is a mental illness in which people believe that they suffer from a serious illness in the absence of an actual medical condition. A common example of hypochondriasis is when a person becomes preoccupied with the idea that their headaches are a result of a brain tumor despite multiple normal brain scans. **Body dysmorphic disorder (BDD)** is a psychological disorder characterized by an excessive concern over a physical flaw. **Pain disorder** is a psychological condition marked by chronic pain in one or more areas of the body without neurological or physiological basis.
4. A **dissociative disorder** is characterized by the disruption of memory, awareness, identity, or perception. It often develops as a coping mechanism following a psychologically traumatic experience, usually during childhood. **Dissociative identity disorder (DID)** is a mental disorder characterized by two or more distinct identities that alternatively control a person's behavior. It's as if they have multiple personalities. **Dissociative amnesia** is the severe impairment of memory recall usually following a stressful or traumatic event. The disorder is characterized by an inability to remember information about one's self and past experiences, which can last on the order of minutes to years. **Depersonalization disorder** is marked by the recurrent detachment from one's self or surroundings. It is characterized by a feeling that you are watching yourself from outside your body or that the things around you are not real.

5. A **personality disorder** is a persistent, inflexible, and maladaptive pattern of behavior that deviates dramatically from what is culturally expected. There are ten primary types of personality disorders that are grouped into three clusters. **Cluster A**—odd and eccentric—includes paranoid, schizoid, and schizotypal personality disorders. **Paranoid personality disorder** is characterized by a pervasive suspicion and distrust of others; patients often perceive being deceived and exploited, even by their friends and family. **Schizoid personality disorder** is characterized by a deeply detached, unemotional, and introspective lifestyle. Schizoid patients are often apathetic with few close friends and poor social skills. **Schizotypal personality disorder** is typically characterized by eccentric behaviors and perceptions, social anxiety, and superstitious beliefs. **Cluster B**—dramatic and emotional—includes antisocial, borderline, histrionic, and narcissistic personality disorders. **Antisocial personality disorder** is far more common in males than females and is characterized by disregard for others, impulsive and aggressive behavior, and a marked lack of conscience. Serial killers and animal abusers often display antisocial personalities. **Borderline personality disorder** is roughly twice as common in females as in males and is characterized by unstable moods, behaviors, self-image, and emotions. Borderline patients often have dramatic and unstable relationships, as well as a strong fear of abandonment. People with **histrionic personality disorder** constantly seek attention, are easily susceptible to peer pressure, and possess dramatic, but shallow and rapidly changing emotions. **Narcissistic personality disorder** is characterized by a lack of empathy, sense of entitlement, and inflated sense of self-importance. **Cluster C**—anxious and fearful—includes avoidant, dependent, and obsessive-compulsive personality disorders. **Avoidant personality disorder** is typically characterized by extreme shyness, sensitivity to criticism and rejection, and strong feelings of inadequacy. **Dependent personality disorder** is characterized by a lack of self-confidence and a persistent need for reassurance and encouragement. People with dependent personalities are often indecisive and fear being independent. Not to be confused with OCD, **obsessive-compulsive personality disorder** is characterized by extreme preoccupation with rules, order, details, and control. These personality types tend to be very rigid, stubborn, and unable to delegate tasks or responsibilities.

Finally, two important disorders that are often covered on the MCAT: schizophrenia and Parkinson's disease. **Schizophrenia** is a brain disorder characterized by an abnormal interpretation of reality accompanied by disorganized and disturbed thoughts, emotions, and behaviors. Schizophrenia is a chronic condition characterized by a wide range of either positive or negative symptoms. **Positive symptoms** include extra behaviors or feelings that are not experienced by healthy people, such as delusions, hallucinations, disorganized thinking, and disorganized behavior. **Negative symptoms** appear as a lack or reduction of normal behaviors or feelings, and often include affective flattening, anhedonia, avolition, and catatonia. **Affective flattening** involves a reduction in emotional expression and is often characterized by a blank face and monotonous voice. **Anhedonia** describes one's inability to experience pleasure and is closely paired with **avolition**, a reduction of motivation. In more severe cases, schizophrenics experience **catatonia**, a state of immobility and sustained, awkward positioning.

There are four distinct subtypes of schizophrenia that are defined by the most significant and predominant symptoms present: paranoid disorganized, catatonic, and undifferentiated. **Paranoid schizophrenia** is characterized by delusions and hallucinations; delusions are typically of grandeur and persecution, hallucinations are typically auditory. **Disorganized schizophrenia** is characterized by affective flattening as well as disorganized thoughts and behavior. Generally speaking, the disorganized subtype is not associated with delusions or hallucinations. **Catatonic schizophrenia** is characterized by a dramatic reduction in movement to the point of immobility and agitated resistance to changing positions, or a dramatic increase of excessive and purposeless movement. **Undifferentiated schizophrenia** lacks distinct, predominant, or persistent symptoms. The patient may display a variety of less severe symptoms, or symptoms that fluctuate.

It has been found that on average, schizophrenics possess a smaller-than-normal thalamus, which may reveal why they have a decreased ability to filter information and focus attention, and are also more likely to possess a smaller-than-normal cortex and decreased frontal lobe activity, which may explain why they have impaired cognitive and perceptual function. Schizophrenia is also highly correlated with overly abundant *dopamine* activity in the brain. Schizophrenia also seems to be highly heritable (at least to some degree).

Parkinson's disease is a progressive nervous system disorder characterized by the loss of dopaminergic neurons within the substantia nigra of the basal ganglia. The most common symptoms can be easily remembered by the acronym TRAP: tremor at rest, rigidity of muscles, akinesia (loss or impairment of voluntary movement), and posture stooped. A specific region of the basal ganglia has accelerated cell death that results in reduced activity of dopamine-secreting cells. Recall that dopamine is a neurotransmitter responsible for communication between different parts of the basal ganglia. There are also **Lewy bodies** found in the midbrain, which are abnormal accumulations of the protein α -synuclein bound to ubiquitin that is found within damaged dopaminergic neurons. Similar to Alzheimer's, Parkinson's disease is neurodegenerative in nature. **Fig. 6.7** shows a summary of the psychological disorders that we have discussed in this section.

| Psychological Disorder | Estimated Prevalence | General Description of Disorder | Subtypes of the Disorder |
|------------------------|----------------------------------|---|---|
| Anxiety Disorders | 18.1% | Characterized by excessive nervousness, worry, and unease. Involves psychological symptoms (e.g., feelings of panic and fear) and physiological symptoms (e.g., shortness of breath, racing heart, sweaty palms). | 1. Generalized Anxiety Disorder 2. Panic Disorder 3. Phobic Disorder 4. Obsessive-compulsive Disorder 5. Post-traumatic Stress Disorder |
| Depressive Disorders | 9.5% | Characterized by a persistently low mood or cyclically extreme moods | 1. Major Depression 2. Persistent Depressive Disorder 3. Bipolar Disorder |
| Somatoform Disorders | 0.2-2.0% | Characterized by physical symptoms that cannot be attributed to a physical cause substance abuse, or another mental illness. Symptoms are real and not fabricated, despite lack of evidence. | 1. Somatization Disorder 2. Conversion Disorder 3. Hypochondriasis 4. Body Dysmorphic Disorder 5. Pain Disorder |
| Dissociative Disorders | 8.7% | Characterized by the disruption of memory, awareness, identity, or perception, usually following a traumatic event (e.g., abuse) | 1. Dissociative Identity Disorder 2. Dissociative Amnesia 3. Depersonalization Disorder |
| Personality Disorders | 9.1% | Characterized by persistent and inflexible behavior that is maladaptive and deviates from cultural expectations. | 1. Cluster A (Odd and Eccentric) 2. Cluster B (Dramatic and Emotional) 3. Cluster C (Anxious and Fearful) |
| Schizophrenia | 1.1% | Characterized by an abnormal interpretation of reality accompanied by disorganized and disturbed thoughts, emotions, and behaviors that severely impair normal living. | 1. Paranoid 2. Disorganized 3. Catatonic 4. Undifferentiated |
| Parkinson's Disease | 1.6% (Americans over age 60) | Characterized by the loss of dopaminergic neurons within the substantia nigra. T. R. A. P. symptoms. | |
| Alzheimer's Disease | 11.0% (Americans over age 65) | Characterized by memory dysfunction due to the formation of amyloid plaques, neurofibrillary tangles, and a general atrophy of the brain. | |

Figure 116: Personality disorders.

7 Critical Analysis and Reasoning Skills

7.1 MCAT Critical Analysis and Reasoning Skills

The process of elimination is very important on the MCAT. If you have an answer set with two choices that are opposites of each other, then one of them is probably correct. If you have two choices that are very similar, odds are that one of them is correct. If there is one answer choice that seems completely different from the other three, try to ignore it until you have evaluated at least one of the other choices. If one of these three similar choices seems to be at least close to right, then you have cut your choices down to two (that one and the ‘different’ one). But if all three similar choices seem obviously wrong, then you can go back to the one that stands out.

7.2 Passage Analysis

When previewing questions, be careful to NOT preview the answer selections that are written below each question. If you make the mistake of previewing all of the possible answers when you preview the questions, you have put yourself at greater risk for experiencing **false pattern recognition**. That is, some of a question’s most tempting (but wrong) answer choices may be written using key words and phrases that do appear in the passage, but which are used in the passage in connection with something other than what you are being asked about in the particular question.

Study summaries carefully—learn how to make your own when you do your outside readings and practice passages. These argument summaries are extremely helpful for answering any evaluative or **inferential** type of question that might be asked, just as making a mental map of a passage to locate any specific bit of information in the passage is very useful for answering **noninferential** questions.