BIO 264 Anatomy & Physiology I

Table of Contents

BYU-I Books

URL: https://books.byui.edu/bio_264_anatomy_phy_I

(n.d.). *BIO 264 Anatomy & Physiology I*. BYU-I Books. https://books.byui.edu/bio_264_anatomy_phy_I

1.0

MODULE 1: TERMINOLOGY/HOMEOSTASIS

1.1

TERMINOLOGY

Integumentary System: hair, skin, nails, sweat glands

Function: To protect body from damage, control body temperature, impede loss of water, and assist in production of vitamin D.

Skeletal System: bones, related cartilage, joints, and ligaments

Function: Framework of the body that lends support and protection, creates blood cells, permits movement, and provides storage for fat and minerals.

Muscular System: muscles and tendons

Function: Facilitate movement, generate body heat, and sustain posture.

Lymphatic System: lymph nodes, lymphatic vessels, various organs

Function: Eliminates foreign materials from circulation, fights illness, regulates tissue fluid level, and absorbs fatty acids contained in the digestive tract.

Respiratory System: lungs, airways, respiratory muscles

Function: Exchanges molecules of carbon dioxide and oxygen between the external environment and the blood. Also maintains blood pH.

Digestive System: mouth, esophagus, stomach, intestines, and various organs

Function: Mechanical and chemical digestion of food, absorbing nutrients, and expelling waste products from the body.

Nervous System: brain, spinal cord, nerves, and sensory receptors

Function: Sensory perception, exercising control over body movement, cognitive reasoning, and a vast array of physiological processes.

Endocrine System: pituitary gland, thyroid gland, pancreas, and glands

Function: Release of hormones: influences growth and development, metabolism, reproduction, and other physiological processes.

Cardiovascular System: heart, blood vessels, and blood

Function: Utilizes blood as a vehicle: distributes gases, nutrients, hormones, and waste products. Also helps with immune response and body temperature maintenance.

Urinary System: kidneys, urinary bladder, ureters, and urethra

Function: Maintains blood pH, regulates water balance, and expels waste products from blood.

Reproductive System:

Female reproductive system includes ovaries, uterus, mammary glands, vagina, and other structures

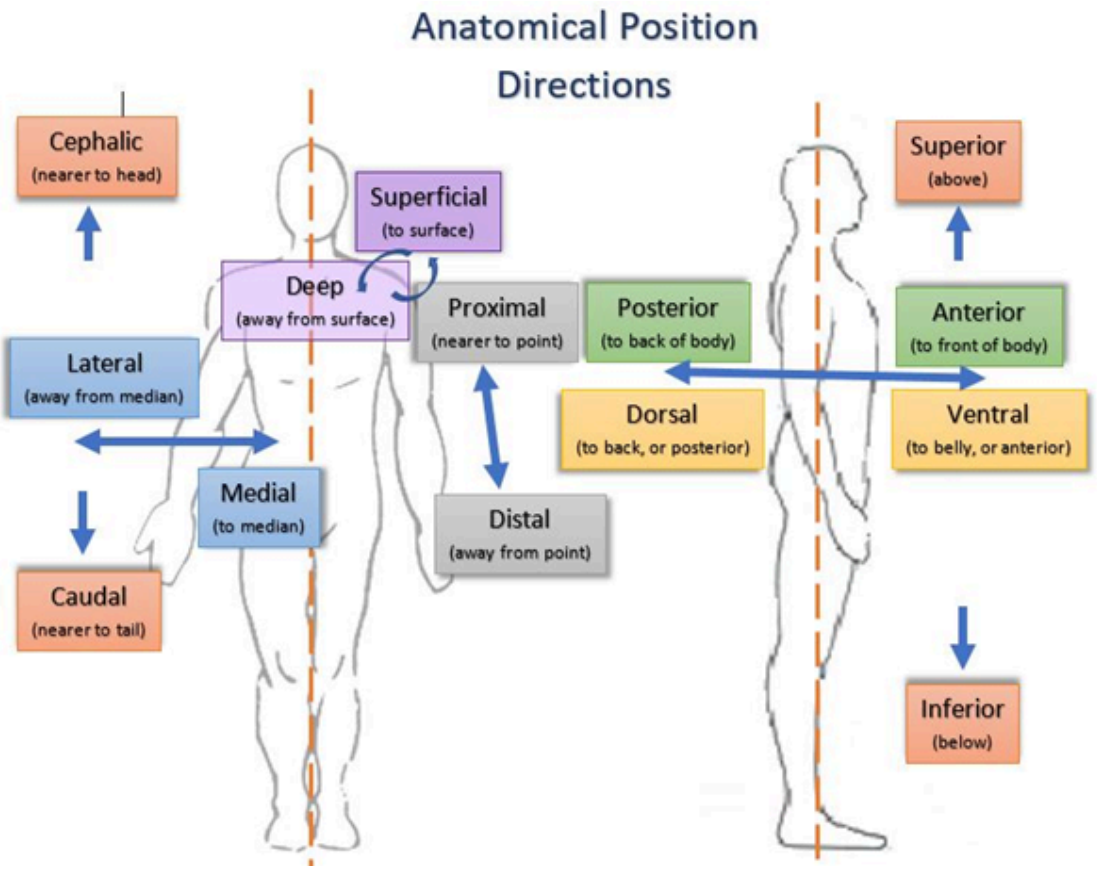
Function: Produces oocytes, provides location for fertilization and fetal development. Also produces hormones which facilitate lactation and sexual behaviors and functions.

Male reproductive system includes the penis, testes, and various other structures

Function: Production of sperm cells and hormones which facilitate sexual behaviors and functions.

1.1.2

Body Directions



Anatomical Position Directions. *Image modified from public domain images of body and directions added by T.Orton 2017*

Superior: above (ex: the head is superior to feet).

Inferior: below (ex: feet are inferior to the head).

Cephalic: relative term meaning nearer to the head (ex: collar bone is cephalic to the sternum).

Caudal: relative term meaning nearer to the tail (ex: sternum is caudal to the collar bone).

Anterior: toward the front of the body (ex: the nose is anterior to the brain).

Posterior: toward the back of the body (ex: the brain is posterior to the nose).

Ventral: toward the belly; equivalent to anterior (ex: the breast is ventral to the spine).

Dorsal: toward the back; equivalent to posterior (ex: the spine is dorsal to the breast).

Proximal: nearer to point of reference or attachment (ex: the shoulder is proximal to the elbow, or the elbow is proximal to the wrist).

Distal: farther away from a point of reference or attachment (ex: the elbow is distal to the shoulder, or the wrist is distal to the elbow.

Lateral: away from the median plane of the body (ex: the shoulder is lateral to the head, or the ear is lateral to the brain).

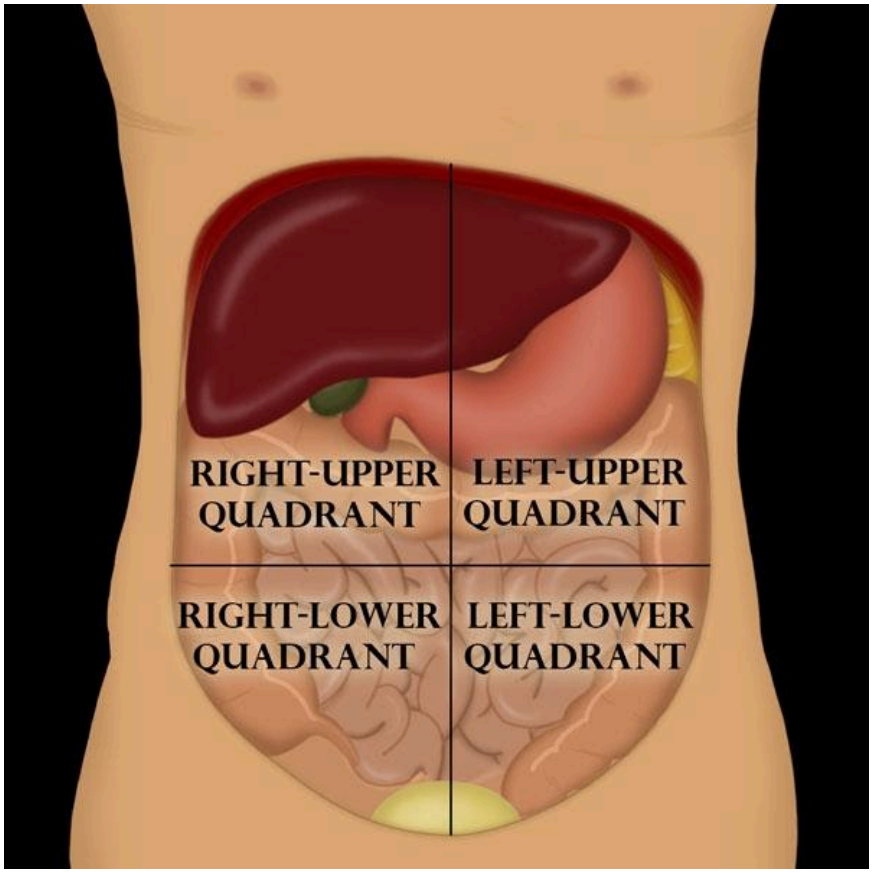
Medial: toward median plane of the body (ex: the head is medial to the shoulder, or the nose is medial to the cheek).

Superficial: toward, at, or pertaining to the surface (ex: the skin is superficial to the muscles).

Deep: away from or below the surface (ex: the muscles are deep to the skin).

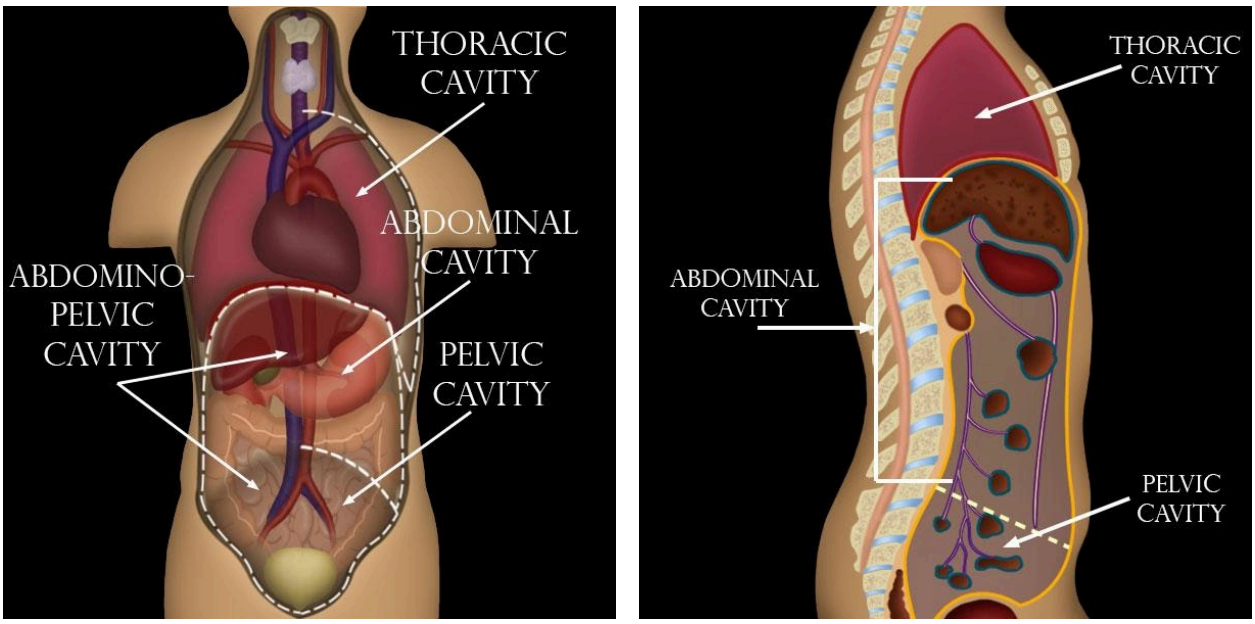
1.1.3

Anatomical Divisions, Subdivisions and Cavities



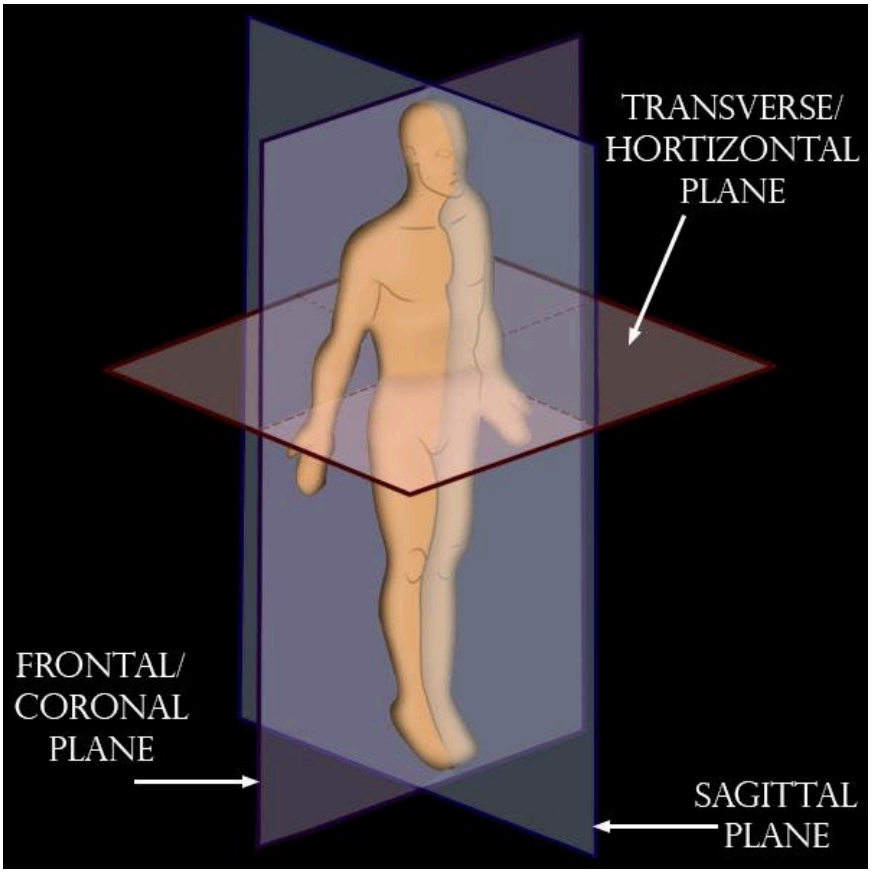
Abdominal Quadrants. *Image by BYU-I student Fall 2013*

The image above shows the quadrants of the abdominal cavity. If you look close, you will find the little extension off the colon called the appendix in the right lower quadrant. Pressure in this area that elicits severe pain is often a sign of inflammation in the appendix.



Cavities of the Body. *Image by BYU-I student Fall 2013*

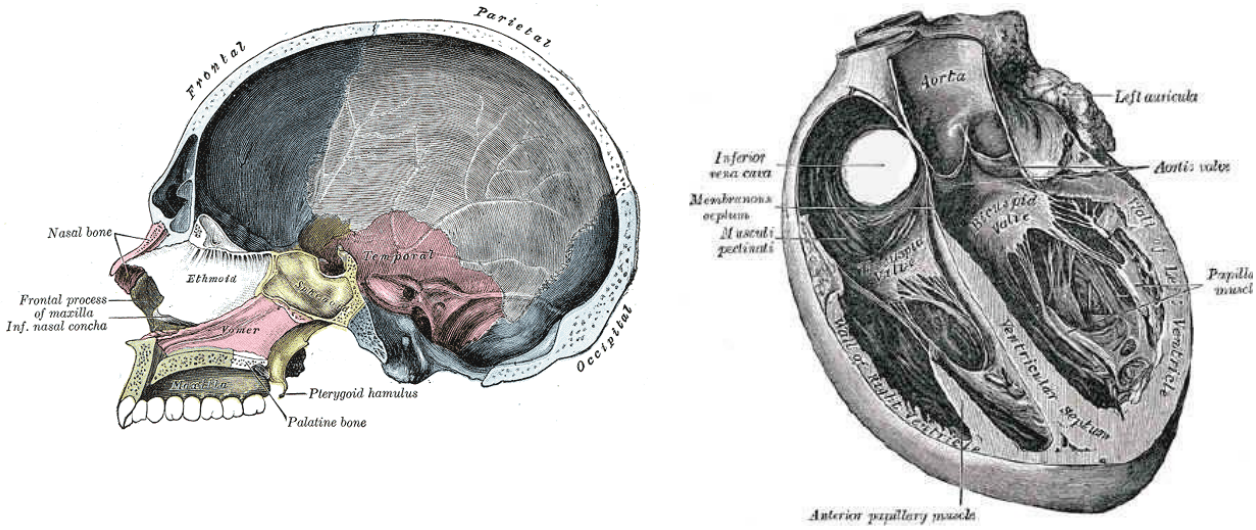
The images above show the “Cavities” of the body. The image on the left is a frontal view, and the image on the right is a mid-sagittal section.



Anatomical Plane Sections. *Image created by a BYU-Idaho student Fall 2013*

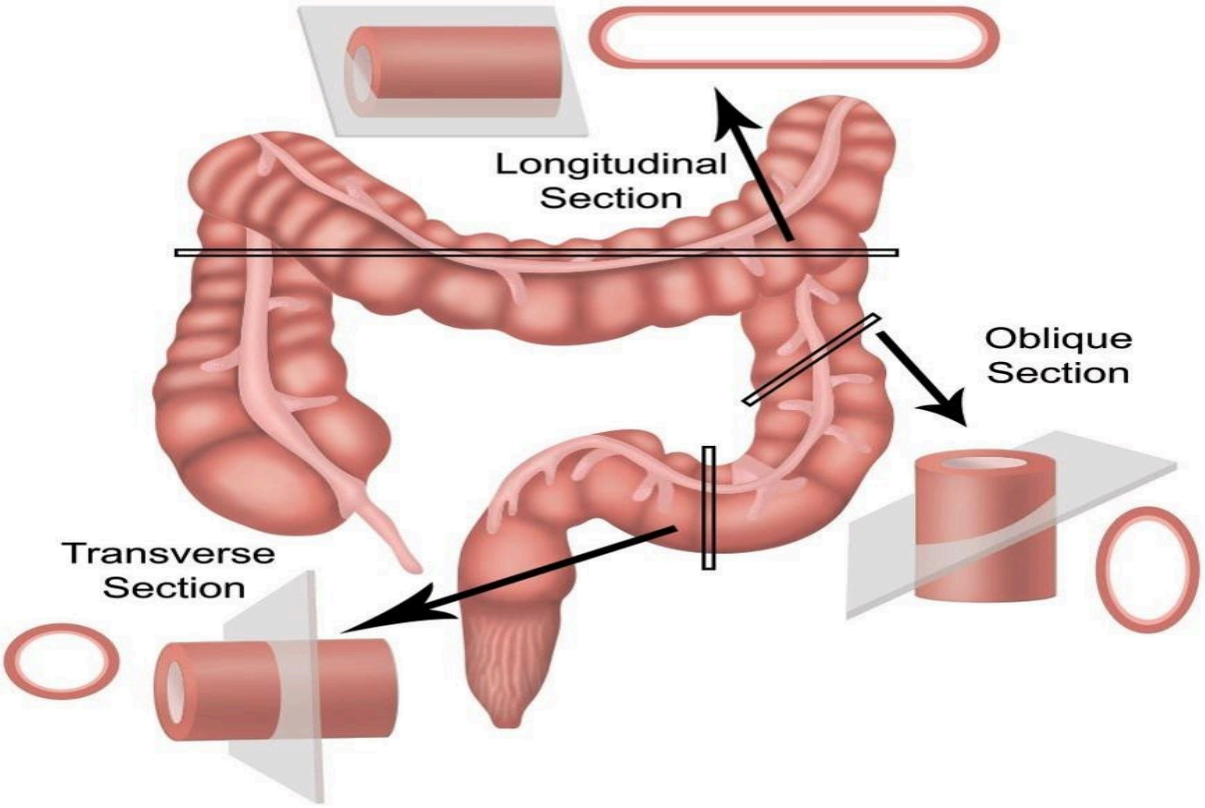
This image shows the possible ways that a “plane” could section the body. The planes are like plexiglass that go through the body in three different orientations. It is a good idea to practice naming the plane that sections a body or body part.

Example: This picture below shows a skull that has been sectioned by the **sagittal plane**. The heart has been sectioned by the **frontal or coronal plane**.



Sectioned View of Skull and Heart. *These images are from Gray’s Anatomy Collection and are Public Domain.*

Sometimes, an organ can be sectioned, and one cannot tell which plane sectioned the organ just by looking at the sectioned image. For example, when an organ wraps around and around in the body, it is not obvious where a single section came from in relation to the body as a whole. This is true with blood vessels and intestines. In these cases, the sections are named as being transverse, longitudinal, or oblique. The image below is an example.



Various Sectioned Views of the Large Intestines. *Image created by BYU-I student Fall 2013*

1.1.4

Prefixes

a- or an-	without or lacking (ex: anorexia)
ab-	away from (ex: abduct)
ad-	toward or adjacent to (ex: adduct)
anti-	against, opposed, or inhibitive (ex: antihistamine)
arthr-	joint (ex: arthroscopy)
auto-	self (ex: autoimmune disease)
bi-	two (ex: biceps)
bio-	life (ex: biology)
carcin-	cancer or tumor (ex: carcinogen)
cardio-	heart (ex: cardiology)
cephal-	head (ex: encephalitis)
cerebro-	brain (ex: CVA or cerebrovascular accident)
chondr-	cartilage (ex: chondrogenesis)
circum-	around or round about (ex: circumcision)
co-	with or together (ex: cofactor)

contra-	against or in opposition to (ex: contraindication)
derm-	skin (ex: dermatologist)
di-	two (ex: disaccharide)
dys-	difficult or abnormal (ex: dyslexia)
ecto-	outside or external (ex: ectopic)
endo-	inside or within (ex: endocardium)
epi-	upon or above (ex: epidermis)
erythro-	red (ex: erythrocyte)
eu-	well, good, or true (ex: eukaryote)
ex-	out of or outside (ex: excretion)
glyco-	sweet or sugar (ex: glycogenolysis)
hemi-	half (ex: hemipalegia)
hepat-	liver (ex: hepatocyte)
hist-	tissue (ex: histology)
hydro-	water or liquid (ex: hydrocele)
hyper-	in excess (ex: hyperglycemia)
hypo-	below normal or under (ex: hypoglycemia)
infra-	beneath or underneath (ex: infrapatellar)
inter-	between (ex: interosseous)
intra-	within or inside of (ex: intramuscular)
iso-	equal (ex: isotope)

leuko-	white (ex: leukocytosis)
macro-	large (ex: macrophage)
melano-	black (ex: melanoma)
micro-	small (ex: microscope)
mono-	one (ex: monosaccharide)
multi-	many or multiple (ex: multicellular)
myo-	muscle (ex: myocardial infarction)
neo-	new (ex: neovascularization)
nephro-	kidney (ex: cystic nephron)
neuro-	nerve (ex: neuroglia)
oculo-	eye (ex: ocular)
odonto-	tooth (ex: orthodontist)
oligo-	deficient or few (ex: oligochromemia)
opthalm-	eye (ex: ophthalmology)
osteo-	bone (ex: osteoporosis)
oto-	ear (ex: otology)
para-	alongside or beyond (ex: parathyroid glands)
peri-	around (ex: pericardium)
physio-	related to nature, physical or physics (ex: physiologic)
pneumo-	air or gas; pertaining to the lungs (ex: pneumonia)
pod-	foot (ex: podiatrist)

poly-	many or much (ex: polysaccharide)
post-	behind or after (ex: post-traumatic)
pre- or pro-	prior to or in front of (ex: prehypertension)
pseudo-	false (ex: pseudostratified epithelium)
sarco	flesh (ex: sarcomere)
semi-	half or partially (ex: semi-permeable)
sclero-	Hard (ex: scleroderma)
somato-	body (ex: somatic)
steno-	narrow or close (ex: aortic stenosis)
sub-	below or under (ex: subclavian)
sym- or syn-	with or together (ex: symphysis pubis and synchondrosis)
tachy-	fast or rapid (ex: tachycardia)
trans-	across or through (ex: transcutaneous)
viscer-	pertaining to internal organs (ex: visceral fat)

1.1.5

Suffixes

-able	capable of (ex: teachable)
-algia	pain (ex: fibromyalgia)
-blast	bud, usually referring to cells which are primitive in nature (ex: Chondroblast)
-clast	pertaining to breaking or tearing down (ex: osteoclast)
-cyte	cell (ex: hepatocyte)
-duct	to lead or draw (ex: adduct and abduct)
-ectomy	to excise or cut out (ex: cholecystectomy)
-emia	blood (ex: leukemia)
-genesis	origin or production (ex: glycogenesis)
-gram	A drawing (ex: sonogram)
-graph	instrument used to record (ex: angiography)
-itis	inflammation (ex: tonsillitis)
-logy	study of (ex: cardiology)
-lysis	to break down or decompose (ex: glycogenolysis)
-oid	resembling (ex: arachnoid mater)

-oma	tumor (ex: carcinoma)
-pathy	disease or disorder (ex: cardiomyopathy)
-phag	to eat or to feed (ex: phagocytosis)
-phil	to love or have an affinity for (ex: hydrophilic)
-stasis	stop or stable state (ex: homeostasis)
-stomy	surgical procedure in which an artificial opening is established (ex: colostomy)

1.1.6

Abbreviations

ACh	acetylcholine
ADH	antidiuretic hormone
ANS	autonomic nervous system
ATP	adenosine triphosphate
AMP	adenosine monophosphate
cAMP	cyclic adenosine monophosphate
CO2	carbon dioxide
DNA	deoxyribonucleic acid
ECG/EKG	electrocardiogram
FAD	flavin adenine dinucleotide
FADH2	reduced flavin adenine dinucleotide
GMP	guanosine monophosphate
cGMP	cyclic guanosine monophosphate

H+	hydrogen ion (acid)
H2CO3	carbonic acid
HCO3-	bicarbonate ion
H2O2	hydrogen peroxide
HCl	hydrochloric acid
HDL	high-density lipoprotein
HR	heart rate
Kg	kilogram
L	liter
LDL	low-density lipoprotein
mOsm	milliosmole
mV	millivolt
Na+	sodium ion
NaCl	sodium chloride
NAD+	nicotinamide adenine dinucleotide
NADH	reduced nicotinamide adenine dinucleotide
NH3	ammonia
NH4	ammonium

NO	nitric oxide
O2	oxygen
OH-	hydroxide
Pi	inorganic phosphate

1.2

HOMEOSTASIS

Homeostasis Defined
Homeostatic Control Systems
Feedback Response Loop

1.2.1

Homeostasis Defined

One of the defining features of warm-blooded animals, like humans, is the ability to maintain a core body temperature that is different from the environmental temperature. The average human body temperature is 98.6°F (37°C), and the body exerts a fair amount of energy ensuring that this temperature stays relatively constant; we call this temperature (98.6°F) the **set point** for body temperature. Different set points for different systems are found throughout the body. For instance, the set point for glucose (blood sugar) is 85 mg/dl, and the set point for sodium is 142 mmol/L. The body uses a variety of organs and organ systems to help ensure that certain **variables** remain as close to their set point value as possible, or at least within a **normal range**. For example, without the assistance of clothing, the human body has a remarkable capacity for keeping the variable of body temperature between 98°F and 100°F, even when placed in environmental conditions that range from 68°F to 130°F. How does the body stay warm at 68°F and cool at 130°F?

To stay warm, the body can increase metabolism, divert blood flow away from the surface, or cause muscles to shiver. All of these mechanisms generate heat. Of course, we could also use our higher cognitive abilities and put some clothes on. Conversely, to stay cool, the body releases water droplets on the surface of the skin, forming sweat, which acts to dissipate heat as the water evaporates. Perhaps most interesting is that sweating, shivering, and blood flow diversions happen automatically; in other words, we do not consciously control them; they just seem to happen. This automatic property of the human body to regulate variables was observed and defined by Claude Bernard in 1854. Then, in 1926, Walter Cannon named this process **homeostasis**. Homeostasis, like many scientific words, is of Greek origin were homeo means "similar or same," and stasis means "standing still or remaining the same." Homeostasis then, by definition, is the ability of the body to maintain relatively stable internal conditions (internal environment) even though the outside world (external environment) is changing. The internal environment is defined as the fluid that surrounds the cells.

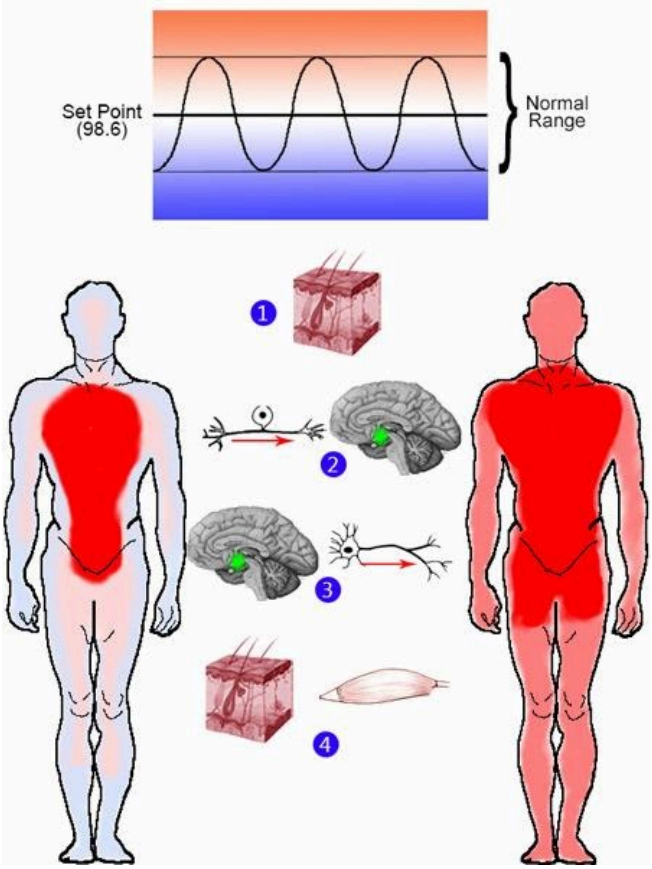
As will be explained, the human body undergoes a multitude of highly complex interactions to maintain homeostasis by ensuring that systems function to hold different variables within their normal ranges. These interactions are essential to the survival of the body. An inability to maintain homeostasis may lead to death or diseases such as: diabetes, dehydration, hyperthermia, and even allergic reactions.

1.2.2

Homeostatic Control Systems

In order to explain how homeostasis works, let's revisit changes that occur to maintain body temperature. How does the body know when to shiver or sweat? The body first needs to detect a temperature change. In the body, this function is attributed to a **receptor**, which is a type of sensor that monitors the environment and detects changes in variables. When conditions cause a change in a variable, we call those conditions **stimuli**. Once a receptor detects a change, it then communicates this change to a **control center**. Control centers are located throughout the body, often in the brain, and are responsible for determining the set point and the appropriate course of action to correct deviations from the set point. Control centers dictate a course of action by communicating with **effectors**. An effector provides the means to correct the deviation. In terms of temperature regulation, the control center is located in the hypothalamus, a small region in the brain, and the effectors would include skeletal muscles (shivering), sweat glands (sweating), and blood vessels (constriction and dilation). It is also interesting that the human body can change a set point for a particular variable. This change is generally temporary and beneficial. For example, the set point for body temperature can change to a higher value in response to infections, called a fever. This increase in temperature aids the immune system in eliminating the pathogen. Consider this critical thinking question: does the set point change observed during a fever represent a negative or positive feedback response? The answer is negative, but why?

An essential component of homeostasis is communication. Communication in the body occurs primarily through two systems: the nervous system and the endocrine system. Regardless of the system used, if communication flows toward the control center from the receptor, it is termed an **afferent** pathway. If information flows from the control center to the effector, it is termed an **efferent** pathway. Collectively, the receptor, afferent pathway, control center, efferent pathway, and effector comprise a **homeostatic control system**. Essentially, all organs and tissues of the body are part of homeostatic control systems and perform functions that help maintain the body's internal environment.



Body Temperature Control by Homeostatic Control System. *Image modified from public domain images of brain and skin. Other elements freehand by JS at BYU-I 2013.*

Homeostatic Control System Steps: Body Temperature

***The numbers below explain the numbers in the picture above**

1. Receptors in the skin and the brain can sense temperature.

2. Information about the temperature travels through afferent neurons to the control center. The control center in this story is the hypothalamus (green dot in the brain picture above).

3. The hypothalamus assesses where the temperature is in relationship to set point (98.6°F). The hypothalamus then sends a signal through efferent neurons to the skin and the muscle tissues.

4. The skin and the muscle tissues are effectors. If the control center determines that the temperature of the body is above the set point, then blood vessels in the skin dilate to divert some of the blood closer to the surface of the body, thus releasing heat in sweat and cooling down the body. Sweat glands can draw water from the blood, trapping heat and releasing it as the sweat (water) is evaporated at the surface of the skin. If the control center determines that the temperature of the body is below the set point, then the blood vessels of the skin constrict to decrease the amount of blood moving to the skin, so the warmer blood instead moves toward the core of the body. Also, sweat glands cease producing sweat. Muscles are another effector that can shiver when it is cold to produce heat in the body.

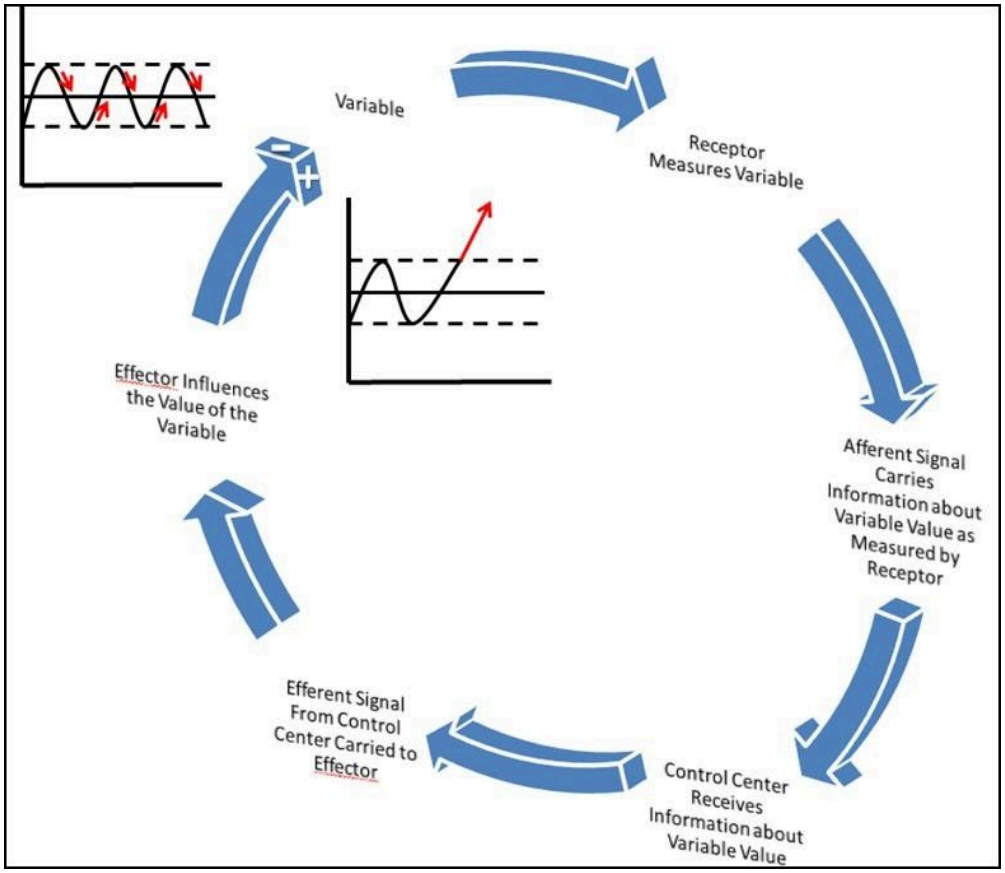
1.2.3

Feedback Response Loop

Homeostatic control systems, like the temperature example above, generally result in Feedback response loops. Feedback response loops start with a stimulus that changes a variable and ends with an effector that changes the variable. If the variable is changed in a way that brings it back towards set point, we call it **negative feedback**. We use the word negative to indicate that the resulting change in the variable is opposite of the initial change. In other words, if a stimulus were to cause the temperature variable to be increased to 99°F, the response of sweating would act to decrease the variable back to 98.6°F. Since the initial stimulus caused an increase in temperature and the resulting response was a decrease in temperature (opposite to the initial change), we call the whole process a negative feedback loop. Regulation of body temperature is only one of many examples of how the body maintains the constancy of the internal environment. Other negative feedback loops that regulate homeostasis include replenishment of oxygen by the lungs, the regulation of the pH of the blood at 7.4, and the regulation of blood glucose by insulin; however, keep in mind that there are many other examples.

Sometimes, the response to a stimulus results in a change to the variable that increases the deviation from the set point. This type of mechanism is called a **positive feedback loop**. Most of the time, positive feedback loops are the result of negative feedback systems that do not adequately correct the problem. For example, in response to a substantial loss of blood, the blood pressure would drop and the negative feedback response would be to increase the heart rate to help return blood pressure to normal. However, if the loss of blood was too great, the increase in heart rate might not be adequate to increase the blood pressure, and as a result, less blood would go to the heart. Since blood carries essential oxygen and nutrients, less blood to the heart would essentially starve the heart. This would result in loss of function and weaker contractions resulting in less blood being pumped, which would result in less blood to the heart and so on. Thus, because the negative feedback response (an increase in heart rate) was not adequate, the end result was that blood pressure continued to drop, causing an increased (positive) deviation from the set point. This situation would require intervention from a medical professional to save the individual.

There are a few examples where positive feedback mechanisms are good. For example, during child birth, labor contractions are enhanced through positive feedback. This is the result of a hormone called oxytocin, which is released from the brain during labor contractions. Oxytocin enters the blood stream from the brain and circulates through the blood to the uterus where it causes more powerful contractions. Contractions of the uterus push the baby's head downward which stretches the cervix. Stretch receptors in the cervix and uterus then send signals to the brain to release more oxytocin, and this positive feedback system continues until birth is accomplished. You may have heard of the drug Pitocin; this is a synthetic form of oxytocin that can be injected into expectant mothers to induce labor or assist contractions when the oxytocin system is not functioning naturally.



Feedback Response Loop. *Image created by JS at BYU-I 2013.*

Above is an image representation of a Feedback Response Loop. Notice that feedback loops can result in Negative or Positive Feedback. The red arrows in the top left graph always show what would happen if the effector(s) always caused the variable to come back to set point (Negative Feedback). The red arrow in the right-hand graph (inside the cycle) shows what would happen if the effector(s) always caused the variable to go further and further from the set point (Positive Feedback).

2.0

MODULE 2: INORGANIC CHEMISTRY

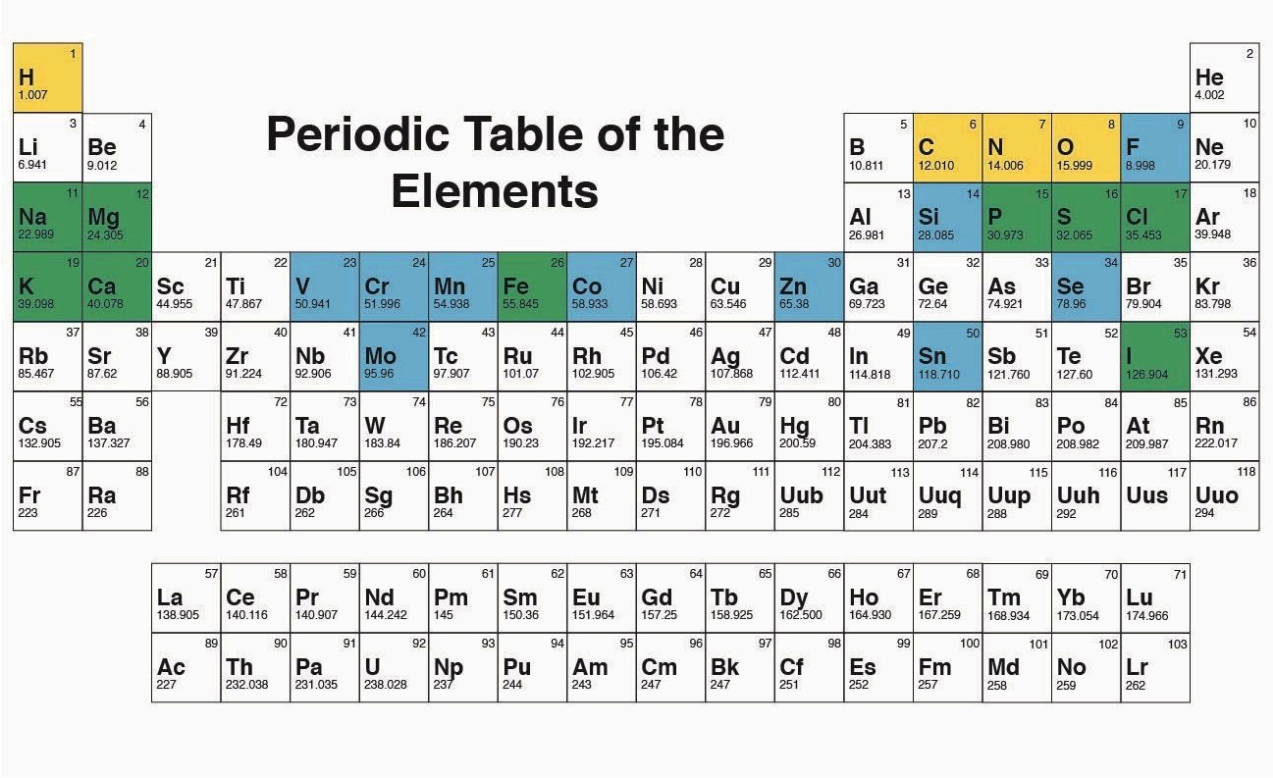
2.1

MATTER

Have you ever wondered why water is a liquid, oxygen is a gas, and sugar is a solid? Or, have you wondered why chlorine gas will kill you if you breathe it but when chlorine is combined with sodium, it is necessary for normal body function and we would die without it? What about fats? Why won't they dissolve in water while salt and sugar do? In the following reading, we will try to answer these and other questions dealing with the chemicals that make up our bodies.

All living and non-living things are composed of **matter**. Using one of its most simple definitions, "Matter is anything that occupies space and has **mass**". Mass is simply the amount of matter that an object contains. Oft times, mass and weight are confused. Weight is a function of gravity pulling on matter. For example, your body is composed of a certain quantity of matter that, when acted upon by gravity, results in your weight of say 150 pounds. If you took that same amount of matter to the moon where the gravitational force is only 1/6 of that of the earth, you would only weigh 25 pounds.

Matter is composed of **elements**. Chemists define elements as "substances that cannot be broken down into simpler materials by chemical processes". Some common elements that you have probably heard of are carbon, hydrogen, oxygen, and nitrogen. The building blocks for elements are atoms, which we will discuss in more detail later. In nature, there are 92 naturally occurring elements. In addition to these, a number have been artificially produced. Based on their chemical properties, these elements can be organized into what is referred to as the **periodic table of the elements**. We will refer back to this table frequently as we discuss the basic chemistry of the elements.



Periodic Table of the Elements, *created by BYU-I student Hannah Crowder, Spring 2011*
The figure above is a "Periodic Table of the Elements." The elements highlighted in yellow make up 96% of our body weight. The nine elements highlighted in green, along with those in yellow, are considered major essential elements. The elements highlighted in blue are considered minor essential elements and are required only in trace amounts in the body. Notice that each element is represented by a 1 or 2 letter symbol. Often, these symbols are the first letter or letters in the name of the element: **H** for hydrogen, C for carbon, and **He** for helium. Occasionally, however, the symbols represent the Latin name for the element; hence, the symbol for sodium is **Na** for the Latin Natrium, and the symbol for Potassium is **K** for the Latin Kalium.

Of the 92 naturally occurring elements, four make up roughly 96% of our body weight, namely **Carbon** (C), **Hydrogen** (H), **Oxygen** (O) and **Nitrogen** (N) (Figure above, yellow highlight). In addition to these four, there are a number of other important, but less abundant, elements found in the body. These include Phosphorus (P), Sodium (Na), Potassium (K), Calcium (Ca), Magnesium (Mg), Sulfur (S), Chlorine (Cl), Iron (Fe), and Iodine (I) (Figure above, green highlight). Having a solid understanding of the elements highlighted in yellow and green will be important as we continue in the study of the human body and cell processes. The elements that are required in trace amounts for normal functioning include Vanadium (V), Chromium (Cr), Manganese (Mn), Cobalt (Co), Molybdenum (Mo), Zinc (Zn), Silicon (Si), Fluorine (F), Selenium (Se) and Tin (Sn) (Figure above, blue highlight).

Subatomic Particles

Electron Configurations

Chemical Bonds

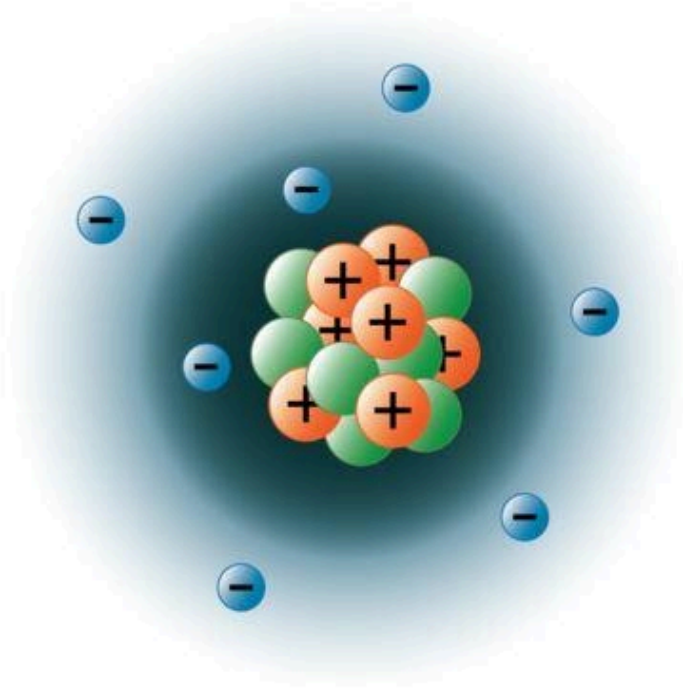
2.1.1

Subatomic Particles

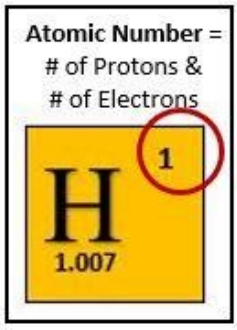
All elements are composed of extremely small particles of matter called **atoms**. We can define an atom as the simplest particle of an element that has the chemical properties of that element. Chemical properties include the physical state of the element (gas, liquid, or solid), the types of bonds the element can form, how it reacts with other elements, etc. Therefore, all of the atoms that make up the element carbon have the same chemical properties.

Physicists have succeeded in blasting atoms apart into dozens of different sub-atomic particles; however, only three of them are stable. These stable sub-atomic particles are the **protons**, **neutrons**, and **electrons**. Protons are positively charged particles, have mass, and are located in the center, or nucleus, of the atom. Neutrons have no charge, have mass, and are also located in the nucleus of the atom. Neutrons bind with protons in a way that helps stabilize the nucleus. Too many or too few neutrons may result in an atomic nucleus that is unstable and may decay to form other elements. We refer to these types of atoms as being radioactive. Although the mass of the neutron is slightly greater than that of a proton, we can assign both of them the relative mass of 1 (1 atomic mass unit or amu). Neutrons and protons constitute almost all of an atom's mass. The third type of stable particle is the electron. Electrons have a negative charge but are extremely small and have a mass only 1/1850 that of a proton or neutron. They are so small that for practical purposes, they do not contribute to the mass of the atom. Electrons move around the nucleus at tremendously high speeds, actually traveling at near the speed of light. Although we often describe the electrons as residing in orbits that circle the nucleus, like planets orbiting the sun, modern physics teaches us that this model is incorrect. These "orbitals" are actually areas in space around the nucleus where the electrons will be located most of the time. This area is often referred to as the electron "cloud." True, it is still a specific area, but it is a bit more amorphous than a spherical orbit. For simplicity, however, we often think of these as satellite-like circular orbitals. The image below represents our current model of a nitrogen atom.

The nitrogen nucleus contains 7 protons (orange) and 7 neutrons (green). The shaded areas around the nucleus represent the electron orbitals (clouds). Electrons (blue) will be found somewhere within these orbitals. (Note: the image is not drawn to scale. It has been suggested that if the nucleus were the size of a basketball, the electrons would be about six kilometers or 3¾ miles away!)



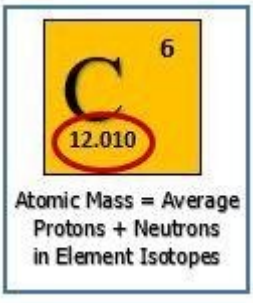
Nitrogen Atom. *Image created by BYU-I student Hannah Crowder Fall 2013*



Atomic Number

Take a look at the periodic table again and notice the number at the top of each box. This number is the **atomic number** for the element and isunique for each different element. For example, the atomic number for hydrogen is 1 indicating it has 1 proton. No other element has an atomic number of 1. For carbon, the atomic number is 6 and, again, no other element has an atomic number of 6. The significance of the atomic number is that it tells us the number of protons in the nucleus of each element. Therefore, all hydrogen atoms have 1 proton (which is why the term hydrogen ion is sometimes used as a substitute for the word proton), and all carbon atoms have 6 protons. In addition, since atoms have a neutral charge, the atomic number also tells us the number of electrons in the atom. In chemical notation, the atomic number for an element is expressed as a subscript preceding the symbol for the element. For example, carbon would be expressed as 6C.

Mass Number (Atomic Mass)

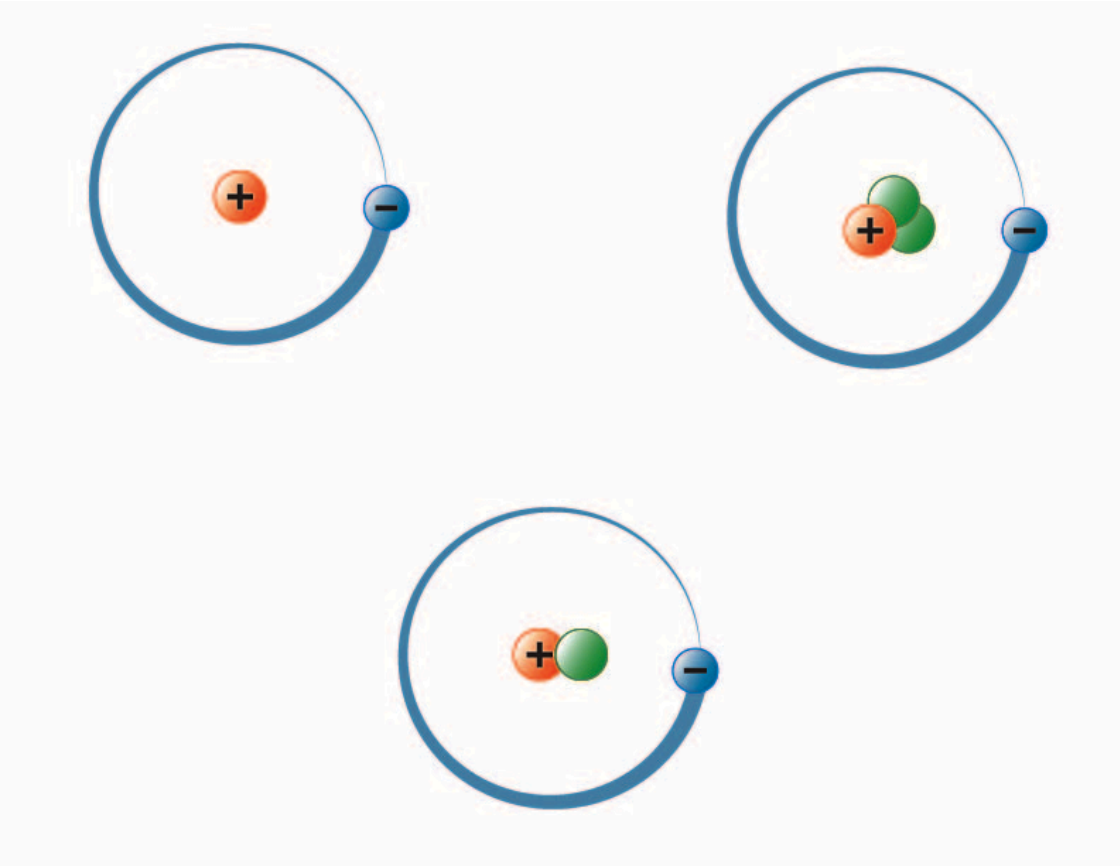


The **mass number** of an atom, as the name implies, tells the total mass of the atom. Since the mass of an electron is extremely small (negligible), it isn't used in the computation of the mass number.Also, recall that the mass of each proton, as well as each neutron, is 1 atomic mass unit. Therefore, the mass number is the sum of the protons and neutrons in the atom.

Since the mass number is the number of protons plus the number of neutrons and the atomic number is the number of protons, you can find the number of neutrons by simply subtracting the atomic number from the mass number. As an example, suppose we have an element with an atomic number of 8 and a mass number of 17. From this information, you can deduce that this element has 8 protons, 8 electrons, and 9 neutrons (17-8=9).

Now, let me throw you a curve ball. As mentioned above, all atoms of a given element have the same number of protons (atomic number). However, different atoms of a given element may have different numbers of neutrons. We say that these are different **isotopes** of the element. For example, there are three isotopes of hydrogen. The most common isotope comprising 99.98% of all hydrogen atoms has a mass number of 1. It, therefore, is composed of one proton, no neutrons, and one electron. The other less abundant isotopes of hydrogen have mass numbers of 2 and 3, respectively. These isotopes differ in the number of neutrons in their nuclei, but all three have one proton and one electron. In reality, there are naturally occurring isotopes of every element, each having its own unique mass number. In chemical notation, the mass number for a given isotope is expressed as a superscript preceding the symbol for the element. The three isotopes for hydrogen would be expressed as ^{1}H, ^{2}H, and ^{3}H.

Since each element is composed of several isotopes, one question that arises is "what is the actual mass of a given element?" Again, if you look at the periodic table above, you will notice a number in the bottom of each box. This is the atomic weight for the element. For example, the atomic weight of hydrogen is 1.00794 amu. This number was derived by computing the average mass of the 3 isotopes of hydrogen. For example, suppose we had 10 boys in our class. If we wanted to know the average weight of the boys, we would add their individual weights together and then divide the total by 10. This would give us their average weight. This is essentially how atomic weights are determined. Since ^{1}H is the most abundant isotope of hydrogen, it makes sense that the atomic weight for hydrogen is very close to the atomic mass of ^{1}H.



Isotopes of Hydrogen. *Image created by BYU-I student Hannah Crowder Fall 2013*
The image above represents the three isotopes of hydrogen. The most common (upper left) has one proton and no neutrons in the nucleus. Deuterium (bottom) has one proton and one neutron, and Tritium (upper right) has one proton and two neutrons.

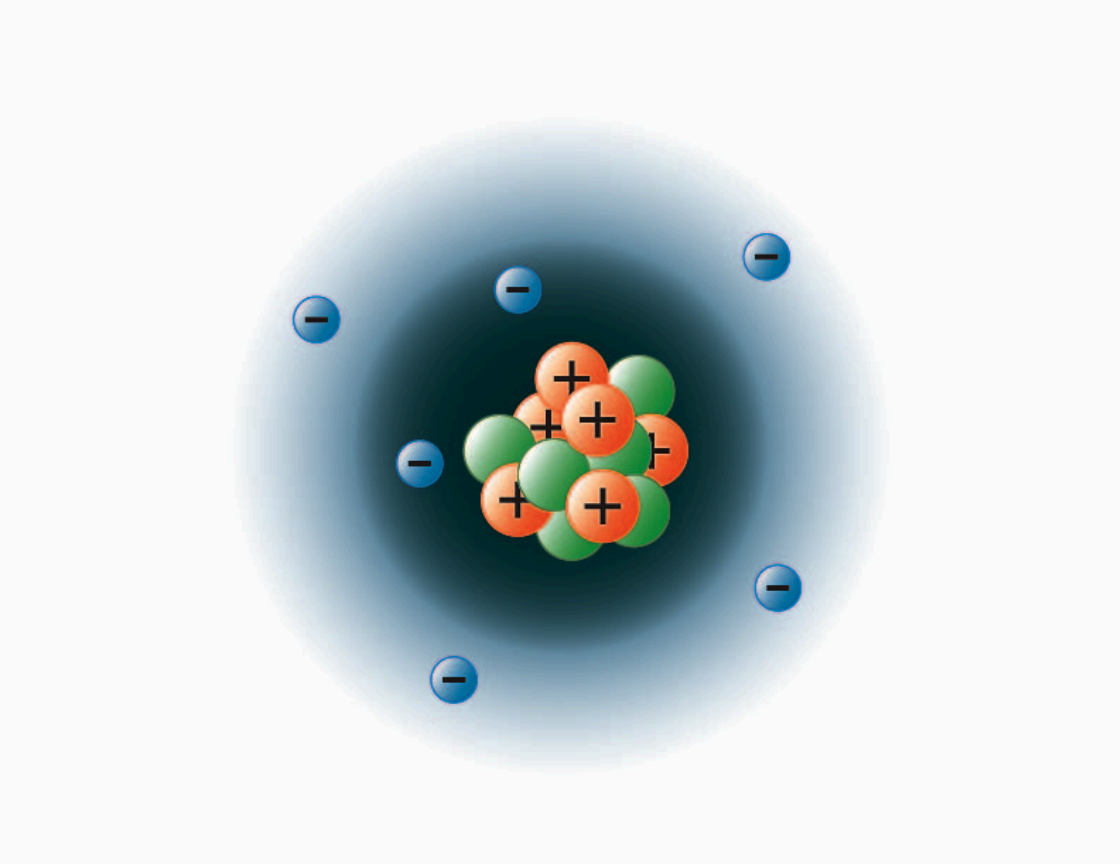
2.1.2

Electron Configurations

The key factor determining the chemical properties of each element is the configuration of its electrons. Likewise, the energy associated with atoms and molecules is a function of their electrons. Think of the atom; it has a positively charged nucleus with negatively charged electrons orbiting the nucleus. Just like the opposite poles of a magnet, oppositely charged particles attract each other. Because of these attractive forces, it requires energy to pull them apart. In an atom, electrons can be moved further away from the nucleus but only if energy of some form is applied. Likewise, when electrons move closer to the nucleus, energy can be released. When I was a kid, we wore fluorescent masks at Halloween. We would shine a light on them to "charge" the mask, and then, when the lights were turned off, they would glow or fluoresce. At the time, I didn't know how they worked; I just knew they were fun. Now, I know that the light, which is electromagnetic radiation, has energy that can be used to push electrons into orbitals further from the nucleus. When the light is turned off, the electrons "fall" back down into a lower orbital, releasing energy which caused the mask to glow in the dark.

As was mentioned above, the electrons of the atom are located in **orbitals**. From our discussion above, we learned that the energy associated with the electrons in an atom is a function of its position or distance from the nucleus. Therefore, electrons in orbitals close to the nucleus possess less energy than electrons in orbitals that are further away from the nucleus. Another important property of orbitals is that each orbital can hold a maximum of 2 electrons. Based on the amount of energy in each orbital, they are arranged into what are referred to as **electron shells** or **energy shells**, which contain one or more orbitals. All of the electrons in a given electron shell have the same amount of energy.

To accommodate the electrons in the largest of the naturally occurring elements, seven electron shells are required. However, most biologically important molecules are considerably smaller, so we will only be dealing with the first three energy shells as we discuss electron configuration. The first shell can only accommodate one orbital; thus, the maximum number of electrons in the first electron shell is two. The second and third shells each contain four orbitals and can, therefore, accommodate eight electrons each. (For those who have or will take more chemistry, I should point out that the third energy shell has more than four orbitals. However, for reasons beyond the scope of this class, we can ignore the other orbitals.) One other important fact is that as electrons are added to electron shells, they occupy the innermost shells first before filling the outer shells. It's like parking spaces at Walmart; those closest to the store fill first, and once they are filled, shoppers have no choice but to park in spaces further away. For example, hydrogen has one electron which is located in the first (innermost) electron shell. Helium has two electrons, both in the first energy shell. All of the space in the first energy shell is now filled. Lithium has three electrons; two of them are in the first shell, and the third electron is in the second electron shell. The reason that this is important to know is because **the chemical properties of an element are determined by the number of electrons in its outer electron shell**. We define the "outer electron shell" as the last shell that has electrons in it, so for hydrogen, its outer electron shell would be the first shell, and for lithium, its outer shell would be the second shell. Each of these elements has one electron in its outer shell, which means that they will have similar chemical properties. Let's try an example. Oxygen has an atomic number of 8, which means it has 8 protons and 8 electrons. How many electrons are there in the outer electron shell of oxygen? The first two electrons will go into the first shell, leaving six to go into the second shell. Therefore, the outer electron shell for oxygen is the second shell, and it has six electrons in it. See if you can determine the number of electrons in the outer shell for sodium, carbon, potassium, chlorine, and neon (see answers at end of section 2.1).



Carbon Atom: *Image created by BYU-I student Hannah Crowder Fall 2013*
The image above represents the electron configuration for carbon. Carbon has an atomic number of 6, hence 6 electrons. The first two electrons fill the first shell (dark blue), and the next four are in the second shell (light blue).

2.1.3

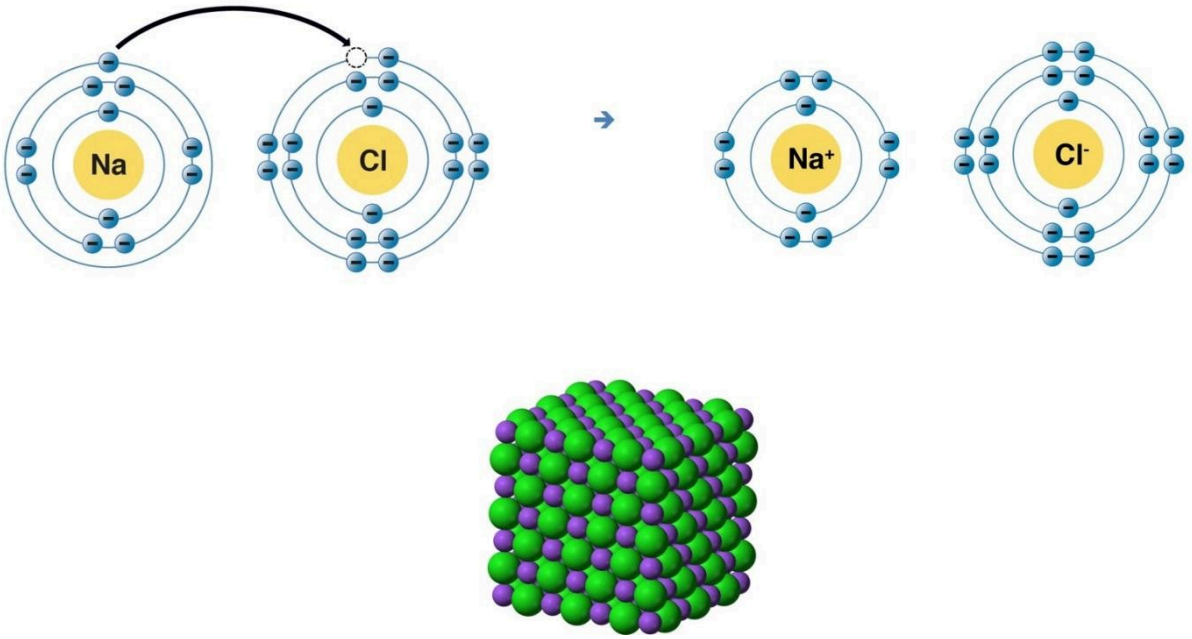
Chemical Bonds

Close examination of the periodic table will show that the atoms of all of the elements in the last column of the table (i.e. helium, neon, argon, etc.) have eight electrons in their outer shells (with the exception of helium). Therefore, their outer electron shells are full. These elements are all gases, and they are all stable, meaning that they do not react with other elements. Chemists often refer to the "rule of 8" which states that if there are eight electrons in the outer electron shell, the element is stable. The atoms of all of the other elements have vacancies in their outer electron shells and will react with other atoms to fill their outer shells. The sections below describe some of the important processes by which atoms become stable. The processes that result in the filling of the outer electron shells result in the formation of chemical bonds. In some cases, this involves the formation of molecules. **Molecules** are two or more atoms held together by the sharing of electrons (described below). Molecules composed of more than one type of element can also be called **compounds**. Hence, H_2 (same element) is a molecule, and H_2O (different elements) is both a molecule and a compound.

Ionic Bonds

To explain how ionic bonds form, we will use common table salt, NaCl, as an example. Sodium has an atomic number of 11; hence, sodium has one electron in its outer electron shell. Chlorine, on the other hand, has an atomic number of 17 and has seven electrons in its outer shell. When these two elements react, sodium gives the one electron in its outer shell to chlorine. Sodium now has eight electrons in its outer shell and is stable. However, the result of losing one electron leaves sodium with one more proton than electron, and therefore, it is now an ion with an electrical charge of +1. An ion is an atom that has a net + or – charge. Ions that have a net positive charge are called **cations**. Chlorine picked up one electron, and now, its outer electron shell is also full and, in the process, has become an ion with a -1 charge (one more electron than proton). Ions that have a net negative charge are called **anions** (think of the term anion as an acronym standing for **a n**egative **ion**). The opposite charges on these ions create an attraction that will hold them together. We refer to this attraction as an **ionic bond**. Figure 3 shows the formation of sodium and chloride ions. By changing the electron configurations of these two elements, their chemical properties have been drastically changed. We require NaCl ($Na^+$and Cl^-) for proper function of our bodies, but both sodium and chlorine with different electron configurations can be lethal. Chlorine gas, Cl_2, is a deadly poison, and elemental sodium (no charge) is a metal that ignites when placed in water. This emphasizes the significance of the statement above that the chemical properties of an element are determined by its electron configuration.

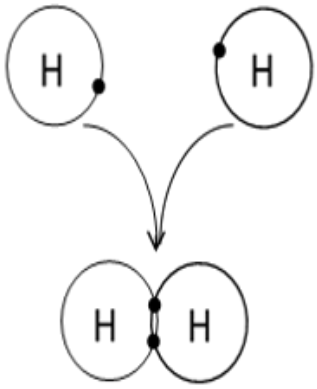
It is also important to note that ionic bonds do not form distinct one-to-one attractions between ions, so technically, ionic bonds do not form molecules. Instead, they form crystalline structures in which each anion is attracted to all of the cations near it, and each cation is attracted to all of the anions near it. Even so, you may still read or hear NaCl being referred to as a molecule/compound.



NaCl Crystal Ionic Bonds: *Image created by BYU-I student Hannah Crowder Fall 2013*
In the image above, the upper left portion represents the formation of an ionic bond. Sodium gives up one electron and becomes a positively charged sodium ion. In the process, its outer electron shell now has eight electrons. Chlorine gains one electron and becomes a negatively charged chloride ion with eight electrons in its outer shell. Upper Right—the negatively charged chloride ions are attracted to the positively charged sodium ions, forming an ionic bond. Bottom—Sodium Chloride crystal. Each sodium ion (purple) is attracted by all of the chloride ions (green) that surround it, and each chloride ion is attracted by all of the sodium ions that surround it.

Covalent Bonds

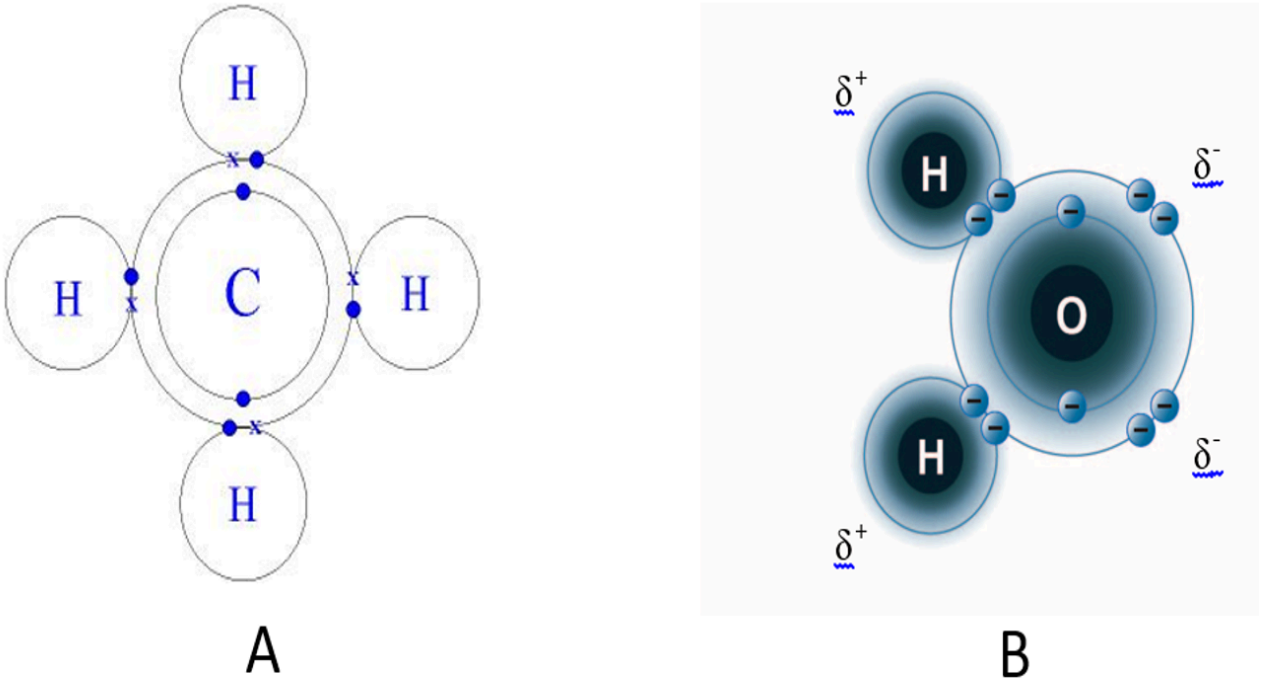
Another way that atoms can fill their outer electron shells is to share electrons. For example, hydrogen has an atomic number of 1 and, therefore, has 1 electron in its outer electron shell. To fill this shell, hydrogen needs one more electron (recall that the first electron shell will hold a maximum of two electrons). One way of filling this shell would be for two hydrogen atoms to unite to form a molecule by sharing electrons with each other. This type of bond, formed by sharing electrons, is called a **covalent bond**. The chemical shorthand for a covalent bond is simply a dash. Therefore, the molecule represented in Figure 6 could be expressed as H-H, with the dash representing the shared electrons. Note that the equal sharing of electrons does not cause either atom to have unequal numbers of electrons and protons; hence, there is no net – or + charge. Also, unlike ionic bonds, which form crystals, covalent bonds create an intimate relationship between the two atoms. That is, these two atoms are linked directly to each other. You could think of ionic bonds as a group date or hanging out and a covalent bond as marriage. The image below represents a covalent bond between two hydrogen atoms.



Hydrogen Covalent Bond: *Image created by MG Fall 2013*

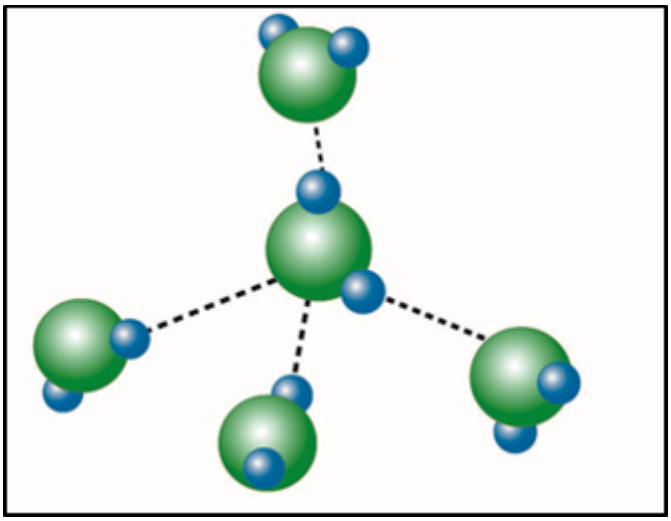
Covalent bonds can also be formed by the sharing of more than one pair of electrons. For example, oxygen has an atomic number of 8 with 6 electrons in its outer electron shell. Two oxygen atoms will combine to form oxygen gas, O_2, by sharing two pairs of electrons, thus completing the outer shell of both oxygen atoms. We refer to this as a **double covalent bond** and represent it by two dashes, O=O, with each dash again representing a pair of shared electrons. It should be noted that triple covalent bonds are also possible by sharing three pairs of electrons. However, in the compounds we will be studying, none have **triple covalent bonds**.

Depending on the atoms involved in the covalent bonds, the electrons can either be shared equally, or the electrons may spend more time with one partner than the other, resulting in unequal sharing of electrons. Two molecules are shown in the figure below. Methane is the image on the left and is a gas that is composed of one carbon and four hydrogen atoms. In this molecule, the electrons are equally shared between the carbon and each hydrogen, forming **non-polar covalent bonds**. The other image on the right is water. In this molecule, the negatively charged electrons are more strongly attracted to the oxygen and, hence, spend more time with the oxygen than with the hydrogen. This creates a molecule that has a slight negative charge at one end (the oxygen end) and a slight positive charge at the other end (the hydrogen end). Since the molecule has oppositely charged ends, we refer to this type of bond as a **polar covalent bond**. Note that the total charge on the molecule is 0, but the ends are charged. Polar covalent bonds and ionic bonds are similar, in that electrons are pulled away from one atom and pushed towards the other. The difference is that with ionic bonds, the electrons are completely removed from one atom, forming the cation, and captured by the other atom, forming the anion.



A: Methane gas – Non-polar covalent bond; B: Water molecule – Polar Covalent Bond. *Image created by BYU-I student Hannah Crowder Fall 2013*

Hydrogen Bonds



Hydrogen Bond: *Image created by BYU-I student Hannah Crowder Fall 2013.*

One other interaction of importance in biological systems is called the **hydrogen bond**. In reality, this is not a bond that forms molecules or ionic crystals; rather, it is an interaction between molecules containing polar covalent bonds. Because of this, it is referred to as an **intermolecular force** or an attraction between two molecules. Hydrogen bonds can only occur between molecules containing polar covalent bonds and are a result of attractions between the oppositely charged ends of these molecules. Note that hydrogen bonds and ionic bonds are similar. The difference is that ionic bonds are created by attractions between oppositely charged ions, while hydrogen bonds are attractions between oppositely charged ends of polar covalent molecules. Although hydrogen bonds are very weak compared to the other bonds we have discussed, they play important roles in many of the compounds we will be studying in this class. For example, most of the important characteristics of water are due to its ability to form hydrogen bonds with

itself and other polar molecules. Likewise, the complex structures of proteins and nucleic acids rely heavily on hydrogen bonding.

Answers to Number of Electrons in Outer Shell Question from 2.1.2:

Sodium (Na): 11 total electrons:

- 2 electrons in 1st energy level,
- 8 electrons in 2nd energy level and
- **1 electron** in the 3rd and outer shell

Carbon (C): 6 total electrons:

- 2 electrons in 1st energy level,
- **4 electrons** in 2nd and outer energy level

Potassium (K): 19 total electrons:

- 2 electrons in 1st energy level,
- 8 electrons in 2nd energy level,
- 8 electrons in the 3rd energy level, and
- **1 electron** in the 4th and outer shell

Chlorine (Cl): 17 total electrons:

- 2 electrons in 1st energy level,
- 8 electrons in 2nd energy level and
- **7 electrons** in the 3rd and outer shell

Neon (Ne): 10 total electrons:

- 2 electrons in 1st energy level,
- **8 electrons** in 2nd and outer energy level

2.2

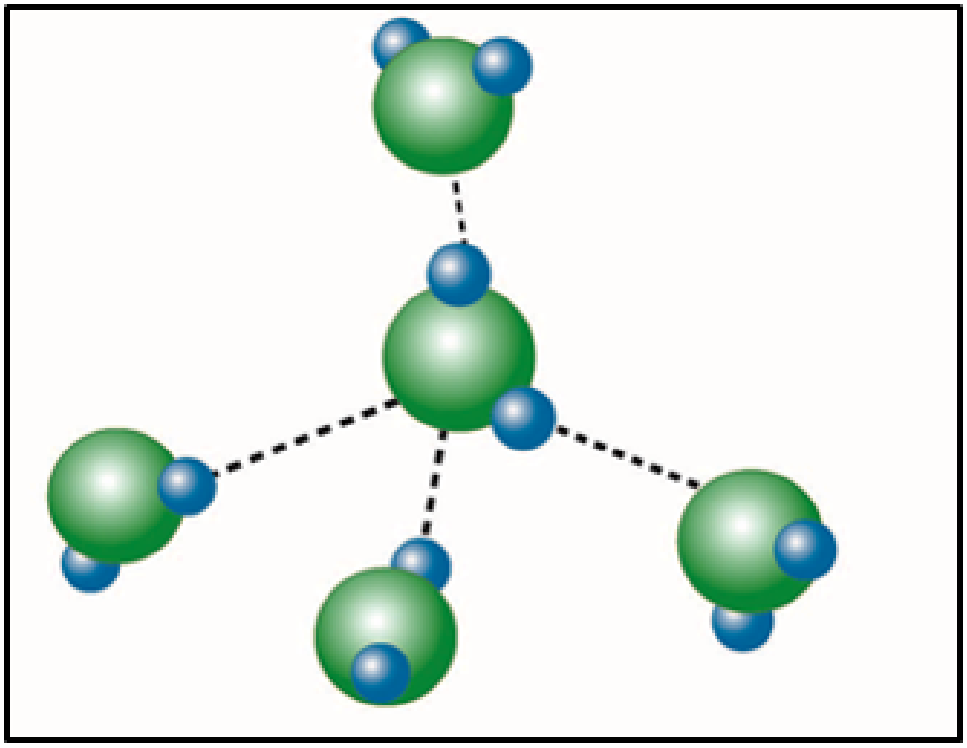
WATER

If someone asked what the most important molecule in your body is, most of us would say water. It makes up about 2/3 of our body weight! That is why a wrestler trying to “make weight” for a match can drop several pounds simply by sweating off water (not a good idea by the way). Whenever there is speculation of life on other planets, the question always arises: Is there water on the planet? What is so special about water? What makes it essential for life? Wouldn’t some other fluid work as well? These are some of the questions that we will attempt to answer in this unit.

2.2.1

Chemical Characteristics of Water

Recall that the water molecule, H_2O, is held together by **polar covalent bonds**. Since the oxygen attracts the electrons in the covalent bonds more strongly than the hydrogen do, the oxygen end of the molecule has a slight negative charge while the hydrogen ends of the molecule have a slight positive charge. Also, recall that molecules composed of polar covalent bonds can participate in weak interactions with other polar molecules through hydrogen bonding. The figure below shows how water molecules form hydrogen bonds with each other. Each water molecule has the potential to form a maximum of four hydrogen bonds with other water molecules. Most of the characteristics of water that we will be talking about are the result of the polar nature of the water molecule and its ability to form hydrogen bonds with itself and other polar molecules. Remember that hydrogen bonds are very weak interactions and can be formed and broken relatively easily. However, as with all bonds, energy is required to break bonds, and energy is released when new bonds are formed. It is the number of these bonds that determine the physical state of the water. For example, in the solid state, each water molecule forms hydrogen bonds with four other molecules, resulting in the formation of a stable, crystal structure known as ice. In the liquid state, each water molecule forms fewer than four bonds (on average 3.4), which are continually rearranging. Water becomes steam when there is enough energy to break all of the hydrogen bonds between water molecules, and they can escape in the form of a gas.



Hydrogen Bonds of Water Molecules: *Image created by BYU-I student Hannah Crowder Fall 2013*

The image above shows hydrogen bonds between water molecules in the solid state. In the liquid state, hydrogen bonds are constantly rearranging (breaking and reforming with other molecules) which allows more movement of the molecules. To our eyes and experience this liquid state can "flow".

Stabilizing Body Temperature

The amount of energy in the form of heat that must be added to or taken from a substance in order to change its temperature is called the heat capacity of the substance. Water has a very high **heat capacity**. In fact, we define the calorie as based on the heat capacity of water. (One calorie is the amount of heat energy necessary to raise the temperature of 1 gram of water 1° Celsius. Note: when reporting the calorie content of food, calorie is written with a capital C. These "big" calories are actually kilocalories or 1000 calories.) Likewise, 1 calorie of energy must be taken away from water to lower the temperature of 1 gram of water by 1° Celsius. Compare this to the heat capacity of air, which is 0.24 calories per gram. This high heat capacity is due to the hydrogen bonds between the water molecules. Temperature is a measure of the total kinetic energy (motion) of a material. Before the water molecules can start moving faster, the hydrogen bonds between the molecules must be broken, which requires the input of energy. Therefore, much of the energy (heat) is used to break the bonds rather than increase the temperature (movement) of the water molecules. By the same token, when heat is removed and the water molecules begin to slow down, new hydrogen bonds form, releasing energy, which helps prevent a big drop in temperature. Since the human body is about 2/3 water, this helps prevent rapid changes in body temperature.

Another property of water that helps stabilize body temperature is its high **heat of vaporization**. This means that in order to convert water from a liquid to a gas, it requires the input of relatively large amounts of energy to increase the movement of the water molecules enough for them to break free from the water molecules around them. As these water molecules move faster and faster, they eventually will have enough energy to completely break away from the liquid and will be converted to a gas (water vapor). When the fastest moving molecules break free, their kinetic energy goes with them, removing heat. This is the basis for the cooling effect of the evaporation of sweat from our skin.

Adhesion, Cohesion and Lubrication

Water is able to stick to other polar substances. This property is referred to as **adhesion**. An excellent example of the importance of this property in the body involves the lungs. A thin layer of water between the outer surface of the lungs and the walls of the thoracic cavity "glues" the lungs to the walls and prevents them from collapsing. **Cohesion** is the sticking together of water molecules. This property prevents the blood from separating as it moves through the blood vessels. Finally, water can act as a lubricant and is found in areas of the body where structures are required to slide past each other. For example, synovial joints (knee, shoulder, ankle, etc.) have a thin layer of water (synovial fluid) between the opposing structures, allowing them to easily slide past one another as the joint moves.

Chemical Reactions

All of the thousands of chemical reactions taking place in our bodies require water. This is because in order to react, the chemicals must be in a watery solution. Water also participates directly in many of the important reactions taking place in the body.

2.2.2

Water and Aqueous Solutions

As mentioned above, before chemicals can react in chemical reactions, they must be in solution. As it turns out, water is an excellent solvent. A **solvent** is a dissolving agent and is the liquid portion of a solution. The molecules that dissolve in the solvent are called the **solutes**. Therefore, in a solution of salt (NaCl) and water, the water is the solvent, and the sodium and chloride are the solutes. A solution in which water is the solvent is called an **aqueous solution**. Although water is an excellent solvent, not everything dissolves readily in water. Materials that dissolve well in water are said to be **hydrophilic** (hydro- = water; -phil- = love), and those that do not dissolve readily are said to be **hydrophobic** (phobia = fear). Usually, if we know the chemical nature of a solute, we can predict how readily it will dissolve in water. For example, compounds that are bound together by ionic bonds tend to be hydrophilic and dissolve readily. The link below shows an animation of an ionic compound dissolving in water. The secret is the ability of the polar water molecules to surround the ions and pull them out of the crystal. When the ions are pulled apart in this manner, we say the compound has become dissociated or ionized. The dissociated ions in the solution are referred to as **electrolytes**. Important electrolytes in our body fluids include Na^+, K^+, Ca^{2+}, Cl^-, HCO_{3^-}, H^+, and Mg^{2+}. These ions participate in many important physiological processes such as nerve impulse conduction, muscle contraction, and regulating water balance.

In addition to ionic compounds, compounds bound together with polar covalent bonds also tend to be hydrophilic. Sucrose, or table sugar ($C_{12}H_{22}O_{11}$), is a good example of a polar compound that readily dissolves in water, forming an aqueous solution.

However, when polar covalent molecules dissolve in water, they do not ionize or separate into smaller particles like ionic compounds do.

Compounds bound together with nonpolar covalent bonds tend to be hydrophobic and do not dissolve readily in water. This is because there are no charged or polar parts to interact with the polar water molecules. Fats and oils are good examples of compounds that are hydrophobic.

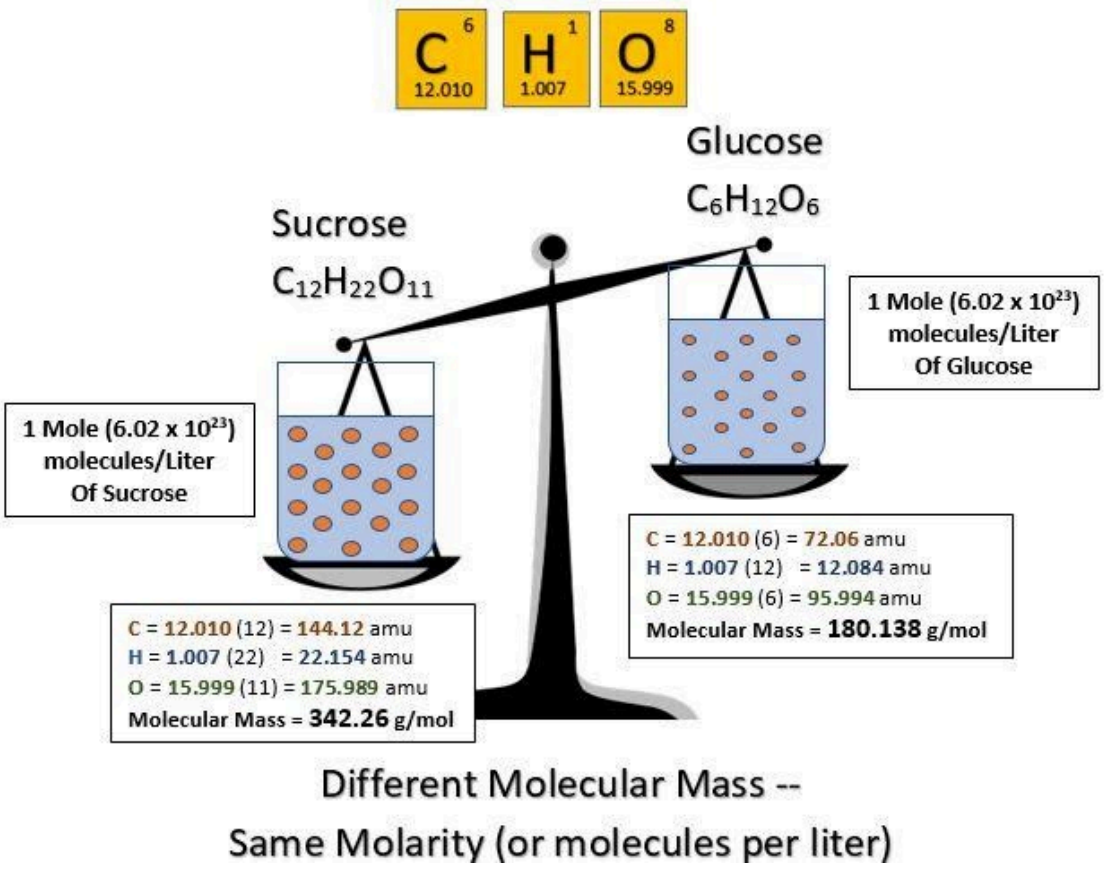
One of the most important structures in the cells of our bodies is the biological membrane. These membranes are stabilized by the hydrophobic and hydrophilic interactions of some special compounds that we will study later.

Solute Concentration

It is often important to know the concentration of the solute in a solution. Some common ways of expressing concentration are straightforward. For example, the normal fasting glucose concentration in the blood is approximately 90 mg glucose per 100 ml of blood. This would be written as 90 mg/dl (dl stands for deciliter or 1/10 of a liter). Another fairly simple method is to express the concentration of the solute in a **percent solution**. This method expresses the concentration as grams of solute per 100 ml of solvent. For example, the normal concentration of NaCl in the blood is 0.9%. This means that there is 0.9 g of NaCl per 100 ml of blood plasma. Both of these methods are fairly easy to visualize. However, they are not very precise. A less obvious but more precise method is to express the concentration as the molarity of the solution. Let's see if we can walk through what this means. First of all, the term *molarity* means **moles** of solute per liter of solution. So, the next question is, what the heck is a mole?

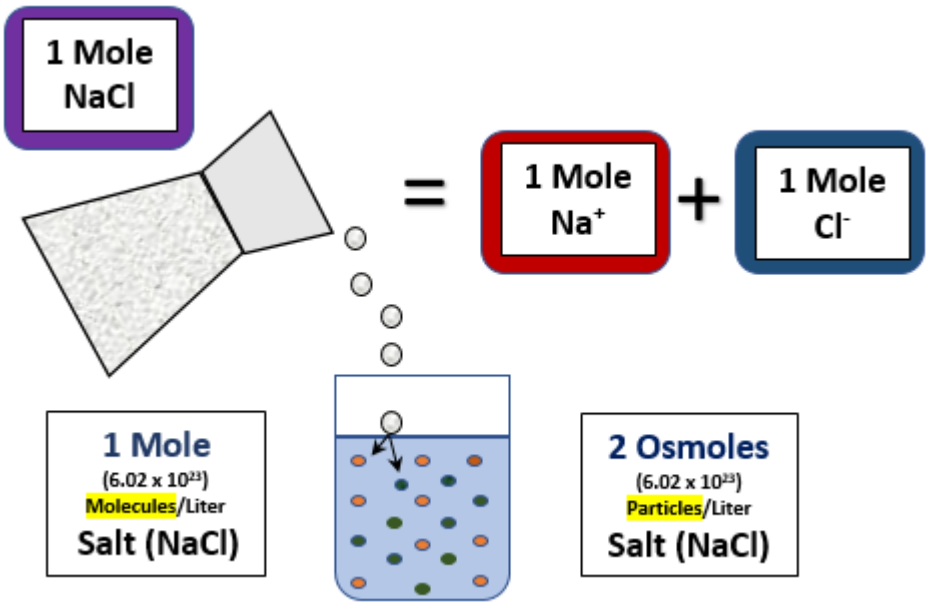
To begin, a mole is a unit of measurement. In order to understand what is meant by “unit of measurement,” let’s start with something you are familiar with, the dozen. We are tempted to ask, “Where in the world did that word come from, and why does it signify the value of 12?” The value of 12 is a unique number because of the early observations of the cycles of the moon, which led to the proposed twelve-month cycle of a year. After a gradual shorting of the Latin word for twelve, “duodecim”, the English derivation of this word became the word dozen—a grouped quantity, signifying the value of twelve. Why group things? For starters, it is easier to go to the store and buy 12 dozen eggs than it is to individually count out 144 eggs. Units of measurements allow us to conveniently talk about large numbers at an understandable level. The concept of unit measurement is absolutely essential when it comes to counting atoms.

Since it is nearly impossible for us to measure 12 atoms, we need to use an understandable unit of measurement. A universally recognized unit of measurement for atoms (or molecules) is the mole (abbreviated mol). Just like one dozen is equivalent to 12, it has been determined that one mole is equivalent to 6.02×10^{23} atoms. The number is called Avogadro's number and is named after the famous scientist, Amedeo Avogadro (1776-1856). Just how big is a mole? If we were to go to a store and buy a bunch of eggs equivalent to the number of molecules in a mole, we could fill the volume of the earth approximately 40 times! Thankfully, molecules are much smaller than eggs; in fact, a mole of sucrose (table sugar) weighs only 342 grams and would barely fill the volume of a tennis ball. The difference, of course, is the size of the atoms that make up the sucrose molecule. So, how did we determine that 342 grams of sucrose contain Avogadro’s number of molecules? First, we need to determine the **molecular mass** of sucrose. To do this, we take the atomic mass of each atom in the compound and add them together. For example, the formula for sucrose is $C_{12}H_{22}O_{11}$. The atomic mass of carbon is 12, and there are 12 carbon atoms in sucrose, so 12 x 12 = 144. The atomic mass of hydrogen is 1, and there are 22 hydrogen atoms in sucrose, so 22 x 1 = 22. Finally, the atomic mass of oxygen is 16, and there are 11 oxygen atoms in sucrose, so 11 x 16 = 176. When we add these together, we get 144 + 22 + 176 = 342. Therefore, the molecular mass of sucrose is 342 (we have rounded the atomic masses to the nearest whole number to make the computations easier). The units for this weight are atomic mass units (amu). Therefore, one sucrose molecule weighs 342 amu. This is much too small of a mass to weigh out on any existing scale (recall that one amu is approximately the mass of a proton or a neutron). However, we can easily weight out 342 g of sucrose. The amu total for a molecule expressed in grams is equal to one mole of that molecule. If we take that amount of sucrose (342 g) and add water until we have one liter of solution, we will have a one molar solution of sucrose. So, in a one molar solution of sucrose, we have 342 g, one mole, or 6.02 x 1023 molecules in one liter of solution. The advantage of expressing concentrations in molarity is that it is an expression that lets us compare the number of molecules in the solutions. For example, a one molar solution of sucrose (molecular mass = 342) and a one molar solution of glucose (molecular mass = 180) will have exactly the same number of molecules per liter of solution even though the molecules are different in size.



Molarity & Molecular Mass: *Image created by T. Orton Summer 2017*

One other expression of concentration that is often useful in physiology is **Osmolarity**, which, as you can probably guess, is **Osmoles** per liter. This is a lot like molarity except that molarity is 6.02 x 10^{23} molecules per liter, whereas osmolarity is 6.02 x 10^{23} particles per liter. At first, it may seem like molecules and particles are the same thing. However, if a solute dissociates (comes apart) when dissolved, you end up with more than one particle per molecule. Take NaCl (salt) for example. When you dissolve one mole of salt in water, each molecule splits in two, so you end up with 2 x (6.02 x 10^{23} particles) in the solution. Hence, a one molar NaCl solution would be a two osmolar NaCl solution (NaCl is a molecule, and Na^+ and Cl^- are particles). This will take on more significance when we talk about osmosis later in the course.



Osmolarity: *Image created by T. Orton Summer 2017*

Questions:

1. How would you prepare a 5% solution of glucose?
2. How would you prepare a 0.5 M solution of glucose? (molecular weight of glucose is 180 amu)
3. What is the osmolarity of a .15 Molar solution of NaCl?
4. How many grams of glucose are there in 0.1 moles of glucose?

Link to answers to Solute Concentration Problems

2.3

ACIDS, BASES, PH AND BUFFERS

Recall that the bonds that bind the oxygen and hydrogen together in water are polar covalent bonds and that covalent compounds typically do not dissociate. However, the polarity of water allows it to form hydrogen bonds with other water molecules in which the negative (oxygen) end of one water molecule is attracted to the positive (hydrogen) end of another water molecule. Although this is a weak attraction, occasionally, the oxygen of one water molecule is able to steal the hydrogen from another water molecule, splitting the water molecules into ions. When this happens, it results in the formation of a **hydrogen ion (H^+)** and a **hydroxide ion (OH^-).** Realize that in pure water, very few water molecules split—about 1 out of every 554,000,000 (who counted?). We can write the equation for this process like this:

$$H_2O \text{ <-- --> } H^+ + OH^-$$

Note that as with all chemical reactions, the reactants and products are in equilibrium, and if that equilibrium is disturbed, the reaction will proceed until a new equilibrium is reached, hence the two-headed arrow in the equation (**<-- -->**).

2.3.1

Acids and Bases

In pure water at 25o C, the concentration of H^+ is always equal to the concentration of OH^-. Both have a concentration of 1.0×10^{-7} Molar. (Placing the symbol for a chemical in brackets $[H^+]$ is chemical shorthand for "concentration of." Therefore, $[H^+]$ is read "the concentration of hydrogen ion.") If we add a substance that results in an increase in $[H^+]$, we say that substance is an **acid**. If we add a substance that results in a decrease in $[H^+]$, we say that substance is a **base**. An acid is any substance that, when added to an aqueous solution, increases the $[H^+]$ of the solution, and a base is any substance that, when added to an aqueous solution, decreases the $[H^+]$ of the solution. A common acid, for example, is hydrochloric acid, HCl. When HCl reacts with water, it dissociates into an H+ and a chloride ion (Cl^-), thus increasing the $[H^+]$. HCl is considered a strong acid because when placed in water, it completely dissociates into its two ions.

$$\mathbf{HCl \text{ --> } H^+ + Cl^-}$$

A weak acid, such as **acetic acid** (CH_3COOH), dissociates into H^+ and CH_3COO^-. However, most remain as acetic acid, and there is a chemical equilibrium between the CH_3COOH and the $H^+ + CH_3COO^-$.

$$\mathbf{CH_3COOH \text{ <-- --> } H^+ + CH_3COO^-}$$

An example of a base is ammonia (NH_3), which will combine with H^+ to form an ammonium ion (NH_4^+), thus removing H^+ from the solution.

$$\mathbf{NH_3 + H^+ \text{ --> } NH_4^+}$$

Another common base is sodium hydroxide (NaOH). How is this a base? When it dissolves, it dissociates into a sodium ion (Na^+) and OH^-, no change in $[H^+]$, right? However, the OH^- will combine with H^+ to form water, thus removing H^+ from the solution.

$$\mathbf{NaOH \text{ --> } Na^+ + OH^-}$$

$$\mathbf{H^+ + OH^- \text{ --> } H_2O}$$

2.3.2

pH

Why do we care about the $[H^+]$ anyway? What is special about this particular ion? Well, it turns out that either too much or too little H^+ can cause serious problems to the body. If the $[H^+]$ is too low, it causes an excitation of the nervous system, resulting in constant contraction of our muscles, including the respiratory muscles, and that is a problem. On the other hand, if the $[H^+]$ is too high, it can result in depression of the nervous system, leading to coma. We use the terms acidic and basic to describe these conditions. If the $[H^+]$ of the solution is greater than 1.0×10^{-7}, we say the solution is **acidic**, and if the $[H^+]$ is less than 1.0×10^{-7}, we say the solution is **basic**.

Because the $[H^+]$ is so important and because it is rather cumbersome to say things like, "the $[H^+]$ of the fluid is 1.0×10^{-7} Molar," chemists have developed a shorthand to express the $[H^+]$. This shorthand expresses the $[H^+]$ as the pH of the solution. The **pH** of a solution is the **negative logarithm of the $[H^+]$** (concentration expressed as moles per liter, M). So, if the $[H^+]$ is 1.0×10^{-7} M, the pH of that solution would be 7 ($-\log 10^{-7}$ is $-(-7)$ or 7). Since this is the pH in which the $[H^+]$ and $[OH^-]$ are equal, we say that this is a **neutral solution**. When using pH, one thing that is a little confusing is that as the $[H^+]$ of a solution goes up, the pH goes down. Suppose that a solution has a $[H^+]$ of 1.0 X 10-6 M. The pH of the solution would be 6, but since the math behind pH is log base 10, the change in pH from 7 to a pH of 6 represents a 10 fold increase in hyrdrogen ions. Moving from a pH of 7 to pH of 5 represents a 100 fold increase. Thus, an **acidic solution** is any solution with a pH<7. Likewise, any solution that has a pH>7 is a **basic solution**. Below is an image that shows the pH of some common solutions.

pH	Examples of solutions
0	Battery acid, strong hydrofluoric acid
1	Hydrochloric acid secreted by stomach lining
2	Lemon juice, gastric acid, vinegar
3	Grapefruit juice, orange juice, soda
4	Tomato juice, acid rain
5	Soft drinking water, black coffee
6	Urine, saliva
7	"Pure" water
8	Sea water
9	Baking soda
10	Great Salt Lake, milk of magnesia
11	Ammonia solution
12	Soapy water
13	Bleach, oven cleaner
14	Liquid drain cleaner

pH Scale and Examples. *Downloaded from Wikimedia Commons Fall 2014; Author: OpenStax College; License: Creative Commons Attribution 3.0 Unported license.*

So, there are two important lessons from this; the lower the pH, the higher the $[H^+]$, and a change in pH of one unit (7 to 6 for example) is a 10-fold change in $[H^+]$. Just for reference, the normal pH of our blood is slightly basic, 7.4 (range = 7.35 – 7.45). If the pH of the blood rises above 7.45, the person is in a state of **alkalosis** (not enough H^+), and if it drops below 7.35, the person is in a state of **acidosis** (too much H^+). In mammals, the pH range of the blood that is considered to be compatible with life is from 6.8 to 7.8. A pH above or below these values usually results in death.

2.3.3

Buffers

Because it is absolutely essential that blood pH is maintained within the narrow range, the body has several mechanisms to regulate the H^+ concentration of the blood. One important defense employed by the body is the various **buffer systems**. Buffers are chemicals that tend to resist changes in pH. Note that buffers do not prevent changes; they resist changes. Let's see if we can figure out how this works.

A typical buffer system is composed of a weak acid and the conjugate base of that acid. Remember, weak acids are those that do not dissociate completely but reach an equilibrium between the reactants and the products of the reaction. An important buffer system in our blood is the bicarbonate buffer system. The components of this system are shown below.

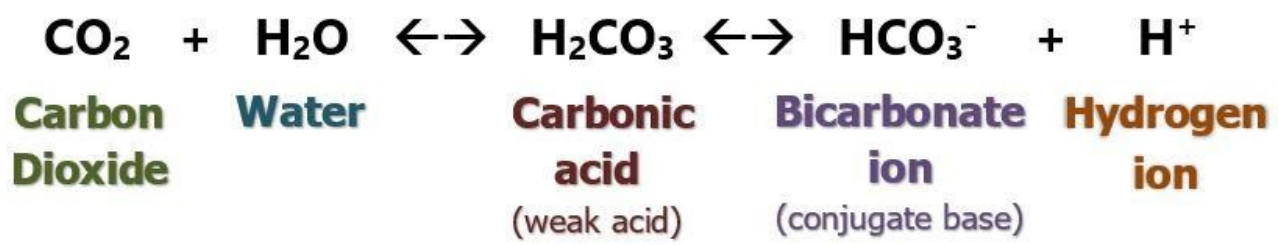
$$H_2CO_3 \text{ <-- --> } H^+ + HCO_3^-$$

Carbonic Acid Hydrogen Ion Bicarbonate Ion

In this case, the carbonic acid is the weak acid, and the bicarbonate ion is its conjugate base. The entire reaction is in equilibrium. If the equilibrium is disrupted by the addition of more hydrogen ions, the reaction will proceed to the left until equilibrium is restored. When it proceeds to the left, some of the excess hydrogen ions will combine with bicarbonate forming carbonic acid, hence removing some of the excess hydrogen ions from the solution. Essentially, the buffer has "soaked up" some of the extra hydrogen ions, thus preventing a large change in pH.

Another way of thinking of this system is to assume it behaves like a teeter-totter. If we have equal weights on each side, the teeter-totter is balanced (in equilibrium). If we add excess weight to one side (excess hydrogen ions), it will be out of balance. The only way to restore balance (equilibrium) is to move some of the excess weight to the opposite side until the teeter-totter is balanced again (equilibrium restored). Obviously, in this simple example, we realize that we cannot move all of the added weight to the opposite side because it would again be out of balance, but if some of the excess weight is moved to the other side, balance can be restored. Like the teeter-totter, when extra hydrogen ions are added, not all can be combined with bicarbonate, so there will still be a few more hydrogen ions than at the beginning (this is why buffers *resist* pH changes instead of *prevent* changes in pH). The pH will decrease, but not nearly as much as it would have if all added hydrogen ions were allowed to remain without being buffered. We could use the same analogy to see what happens when hydrogen ions are removed from the solution by the addition of a base. Since the equation is again out of equilibrium, the reaction will proceed to the right until some of the hydrogen ions have been replaced. Again, there will be a slight increase in pH, but not nearly as great as would happen in the absence of the buffer.

We will be discussing the bicarbonate buffer system in greater depth in upcoming modules. Here is an introduction to the full equation that shifts to maintain balance in our bodies:



3.0

MODULE 3: ORGANIC CHEMISTRY

3.1

CARBOHYDRATES

There are roughly 92 naturally occurring elements on earth, but only four make up about 96% of the mass of the human body: oxygen, carbon, hydrogen and nitrogen. These elements combine to form life-sustaining biomolecules, which can be divided into four major groups: carbohydrates, lipids, proteins, and nucleic acids. Carbohydrates, proteins, and lipids are used by cells as the building blocks for cells and for energy, while nucleic acids are the basis of genetic material (DNA and RNA). Carbohydrates are the most abundant of the biomolecules. If we were to identify the most important carbohydrate molecule on the planet, in terms of its ability to sustain life, we would undoubtedly select the monosaccharide glucose. Without glucose, nearly all animal life as we know it could not exist.

Carbohydrates can be classified into 4 major subtypes: monosaccharides, disaccharides, oligosaccharides and polysaccharides. These classifications are based on both the size and function of the molecule. The name "saccharide" is derived from Greek; it means “sugar.” Monosaccharides are the simplest form of carbohydrates and are composed of a single molecule or subunit. The disaccharides are composed of two monosaccharides linked together, while oligosaccharides are composed of between 3 and 20 monosaccharides and polysaccharides consist of hundreds or thousands of monosaccharides linked together. We will now examine each of these types of carbohydrates.

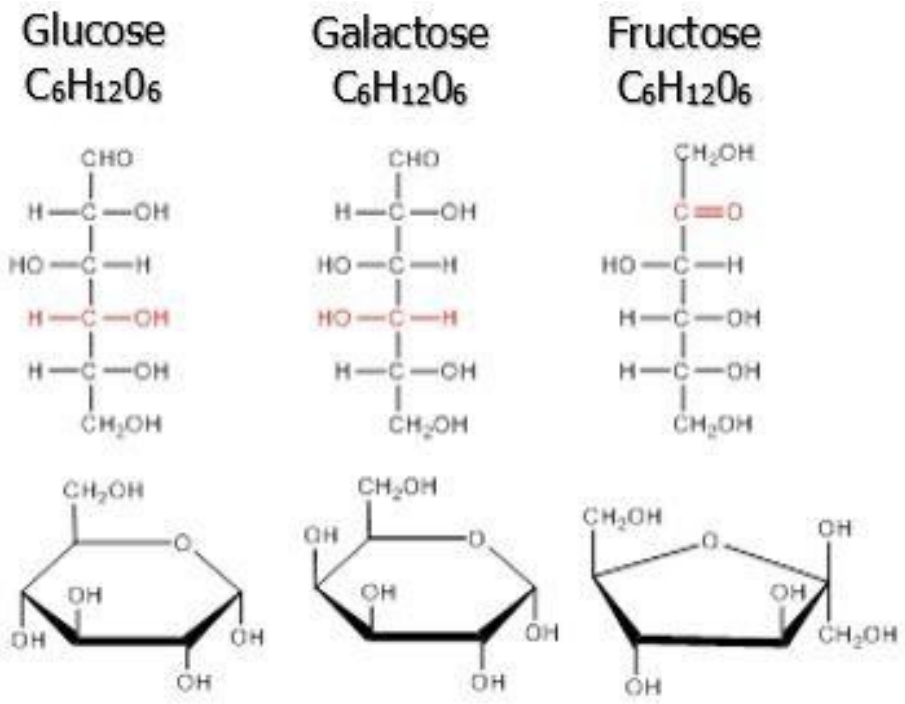
- Monosaccharides
- Disaccharides
- Polysaccharides
- Oligosaccharides

3.1.1

Monosaccharides

Monosaccharides (mono = one, saccharide = sugar) are the basic subunits of carbohydrates. They contain from 3–7 carbons and have the general formula of $(CH_2O)n$ where n ranges from 3–7 (5 or 6 being the most common). For example, if n = 6, the formula for the monosaccharide would be $C_6H_{12}O_6$. Please note that the ratio of carbon to water (H_2O) is 1:1 in a monosaccharide, giving credence to the name carbohydrate. Note also that monosaccharides contain a significant amount of oxygen. Carbohydrates have the highest oxygen to carbon ratio of any of the important organic molecules. These oxygens can increase the solubility of carbohydrates in water (due to the increased number of polar covalent bonds). This is evident in tissues like cartilage, which contains a lot of polysaccharide molecules that act as molecular sponges to hold water, maintaining the cushioning and lubricative functions of certain cartilage types.

Common monosaccharides include **glucose**, **fructose**, **galactose**, **ribose**, **and deoxyribose**. Notice that the name of each of these sugars ends with the suffix -ose. This suffix, -ose, means full, specifically full of oxygen. The names of most sugars will end with this suffix. The structures of three common dietary monosaccharides are shown in the figure below. Note that the molecules can exist in two different forms. When they are in a dry or powdered state, they exist as a linear molecule (top), but when dissolved in water, they adopt a ringed form with oxygen being one of the members of the ring (bottom). Since all of the molecules in our bodies exist as aqueous solutions, the ringed form is how we find monosaccharides in the body. Note also that all three of these compounds have six carbons; hence, they have the same molecular formula, $C_6H_{12}O_6$. However, their structural formulas are different (see figure below). Molecules with the same molecular formula but different structures are called **isomers**.



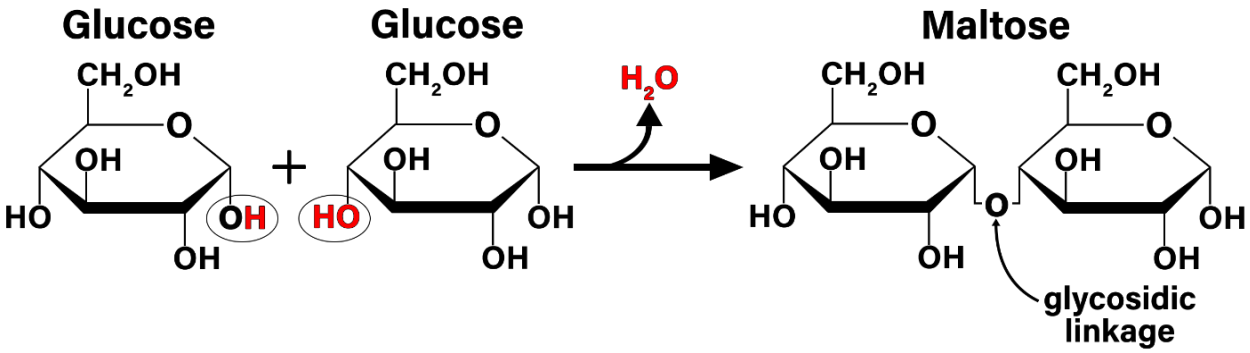
Linear and Ring Structure of Isomers of $C_6H_{12}O_6$. *Image created by MG 2013*

Glucose, also called dextrose, is the predominant sugar in our blood. When we speak of blood sugar levels, we are really talking about blood glucose levels. Our bodies get glucose primarily from the digestion of disaccharides and polysaccharides. Once these carbohydrates are broken down to glucose in the small intestine, the glucose is absorbed into the blood and transported to the various organs of the body. There, it can be metabolized to provide fuel for cellular metabolism. If it is not immediately needed for metabolism it can be stored as glycogen (more about this complex carbohydrate later) in the liver and muscle or converted to triglycerides (fat) and stored in the fat cells. Importantly, in the absence of carbohydrate ingestion (a low-carb diet for example), the body can actually make its own glucose in a process called gluconeogenesis.

3.1.2

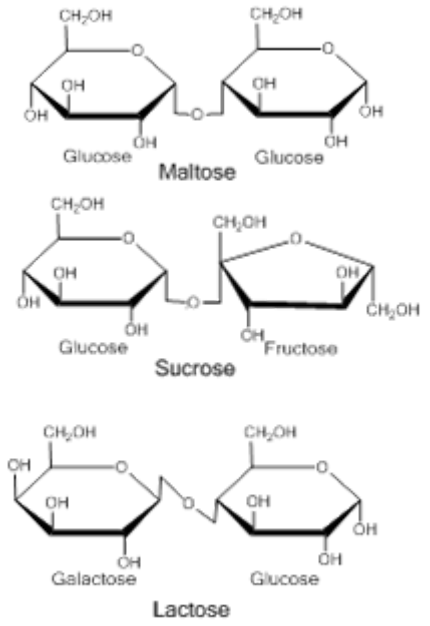
Disaccharides

Disaccharides (Di = two, saccharide = sugar) are formed when two monosaccharide molecules are joined together. This link occurs between -OH functional groups called hydroxyl groups, as shown in the figure below. These groups are joined together by the removal of water. Because a molecule of water is generated, this reaction is called a dehydration synthesis reaction. This is a common type of synthesis reaction that we will see again when we learn about the formation of lipids and proteins.



Dehydration Synthesis Reaction Showing the Formation of Maltose.
Image by BYU-Idaho professor Spring 2021
The image above shows a dehydration synthesis reaction. The reactive hydroxyl groups (-OH) are circled. The hydrogens and oxygen that will be removed to form water are colored red. The resulting linkage is called a glycosidic linkage.

There are three important disaccharides that we will discuss: sucrose, lactose, and maltose. In all three of these disaccharides, glucose is one of the monosaccharides that make them up. The figure below shows the structure of these disaccharides, and the table below outlines their characteristics.



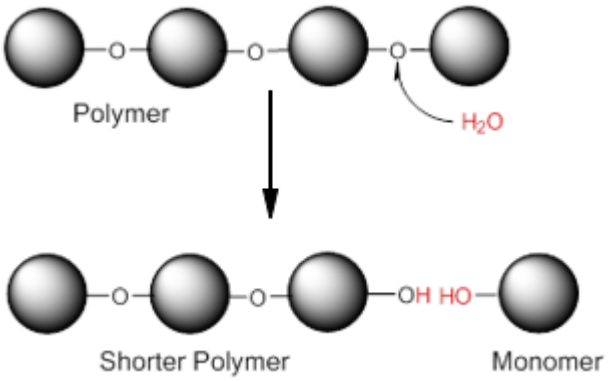
Disaccharide Structure: Image created by MG, 2013

The image above shows the structures of the three common dietary disaccharides. All contain glucose as one of their subunits. The difference between the three is the second subunit.

Table: Characteristics of three common disaccharides.

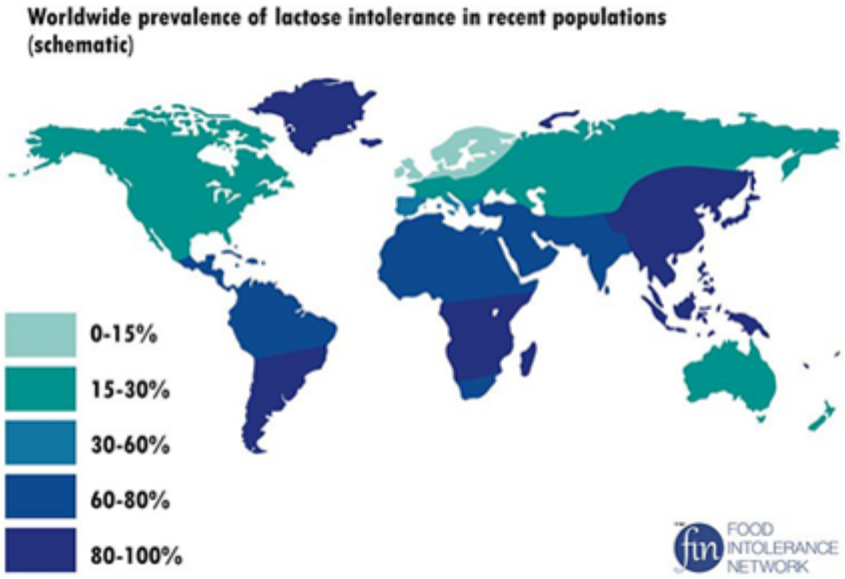
Name	Combined Monosaccharides	Nutritional Information
Sucrose	Glucose + Fructose	The most common dietary disaccharide. Naturally found in beets, cane sugar, brown sugar, maple syrup, and honey. You know it as table sugar.
Lactose	Glucose + Galactose	Found in dairy products. This is the least sweet of the disaccharides.
Maltose	Glucose + Glucose	Found in foods including breakfast cereals, germinating seeds, and beer.

Only monosaccharides can be absorbed from the digestive tract into the blood. Therefore, in order to enter the body, disaccharides must first be broken down (or digested) into their monosaccharide subunits. In the small intestine, there are specific enzymes for each of these disaccharides: sucrase to digest sucrose, lactase to digest lactose, and maltase to digest maltose. The reaction for digestion is essentially the reverse of the dehydration synthesis reaction (i.e. water is added back into the bond to break it). This type of reaction is called a hydrolysis reaction. Because disaccharides are easily digested and quickly absorbed into the blood, they, along with the monosaccharides, are often referred to as the simple sugars.



Hydrolysis Reaction. Image created by BYU-I Student Hannah Crowder, 2013
The image above shows a hydrolysis reaction. Bonds between the monomers in a polymer can be broken by the enzymatic addition of water to the bonds. Monomers can be defined as a single molecule that can bind to other molecules to form a polymer.

What if one of the enzymes that hydrolyze disaccharides is missing? This is actually the case for a vast majority of mammals, including humans. Because most mammals do not consume milk once they are adults, they no longer need the enzyme lactase to digest lactose. Because of this, the body stops making the enzyme. However, if lactose is not broken down into its monosaccharide subunits, it cannot be absorbed and instead passes into the large intestine. The bacteria that live in the large intestine love lactose and start eating it. Unfortunately, when they eat a lot of lactose, they produce a lot of gas. Also, the lactose pulls water into the large intestine by osmosis. Symptoms of lactose intolerance include abdominal bloating, diarrhea, abdominal cramps, flatulence (gas), and nausea. The symptoms are due to undigested lactose moving into the large intestine. Worldwide, about 75% of the adult population experiences some degree of lactose intolerance. However, the incidence differs greatly from country to country (see figure below). Typically, northern Europeans and their descendants have the lowest incidence, mainly due to the fact that in their culture, cattle and goats were domesticated long ago, and the milk products from these animals are an important source of nutrition.



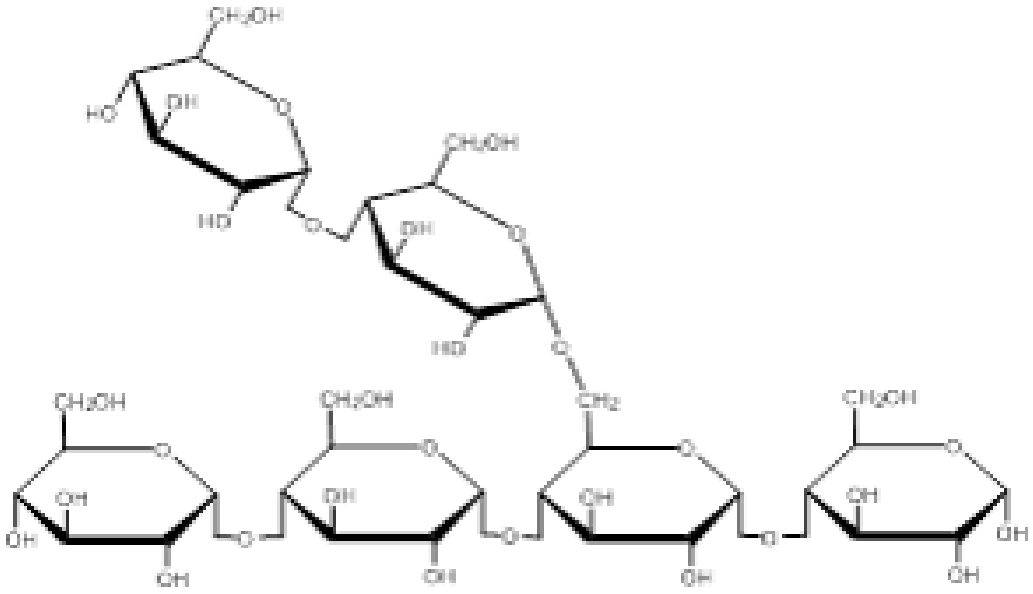
Worldwide Incidence of Lactose Intolerance. Image downloaded from Wikimedia Commons Dec 2013: Author: NmiPortal; Site:

https://commons.wikimedia.org/wiki/File:Worldwide_prevalence_of_lactose_intolerance_in_recent_populations.jpg; License: Creative Commons Attribution-Share Alike 3.0 Unported

3.1.3

Polysaccharides

Polysaccharides are long chains of monosaccharide subunits linked together through dehydration synthesis reactions. Typically these chains contain hundreds to thousands of monosaccharides linked together through glycosidic linkages. Because of their length, polysaccharides are considered complex carbohydrates., Polysaccharides can be classified into two based on their function as either energy storage or anatomical structure. Storage polysaccharides include starch and glycogen. Plants and animals store sugar for energy use in the form of glycogen (animals) and starch (plants). Starch is a large polymer of glucose subunits and may be branched or linear. Amylose is a long, unbranched chain of glucose subunits. Amylopectin, on the other hand, has a branched structure (see figure below). It is the proportion of each form of starch in a particular food that determines the food's ability to be digested. Foods with a large amount of amylopectin are digested and absorbed rapidly because of its many branches, which facilitates hydrolysis. Foods that have higher levels of amylose break down at a slower rate. Some examples of starches include seeds, grains, corn, beans, potatoes, and rice.

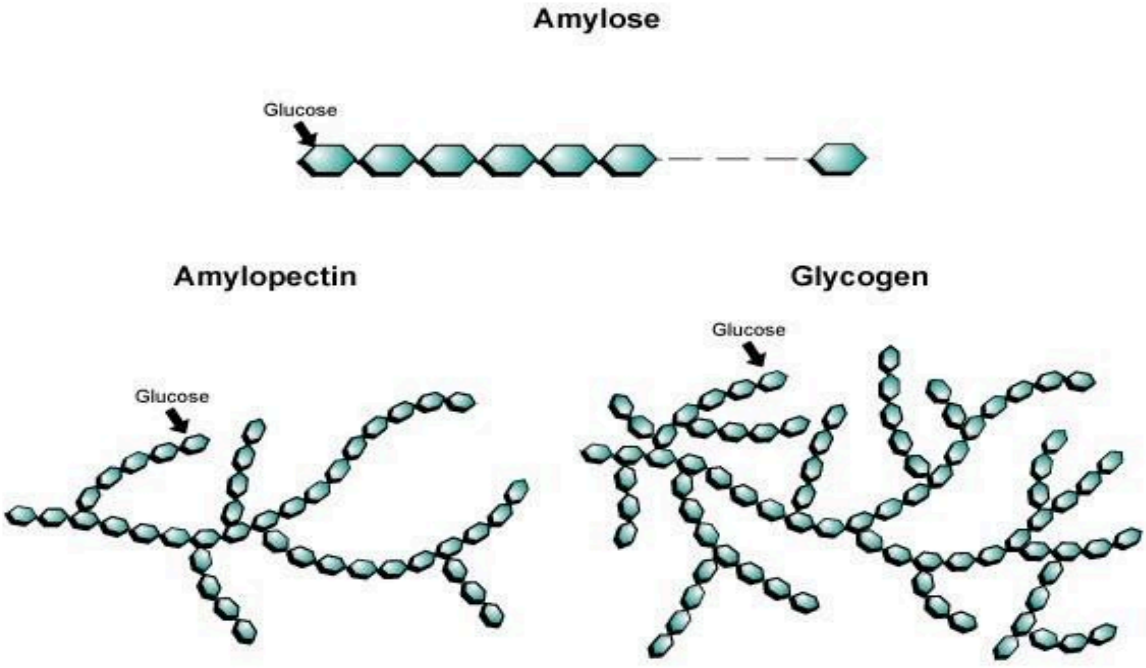


Branched Polysaccharide Amylopectin: Image created by MG, 2013

The image above shows branching in a polysaccharide molecule. Branching allows increased enzymatic breakdown and faster digestion.

Glycogen is the storage form of carbohydrates in animals. Glycogen, like starch, is a polymer of glucose subunits. It is similar in structure to amylopectin, but it is even more highly branched. We store glycogen primarily in our livers and skeletal muscles. The branched structure of glycogen allows for easy breakdown by enzymes in the body to release the glucose, so it can be utilized for energy. Glycogen stored in the muscle provides the energy required by the muscle for exercise, especially during high-intensity and endurance activities. Glycogen stored in the liver is utilized to provide other

tissues with energy, such as the neurons in the nervous system. The glycogen in skeletal muscle can be depleted in as little as one hour of vigorous exercise. On the other hand, during a fast, liver glycogen will last 12–24 hours.



Amylose, Amylopectin & Glycogen Structure. Image created by BYU-I student Hannah Crowder, 2013

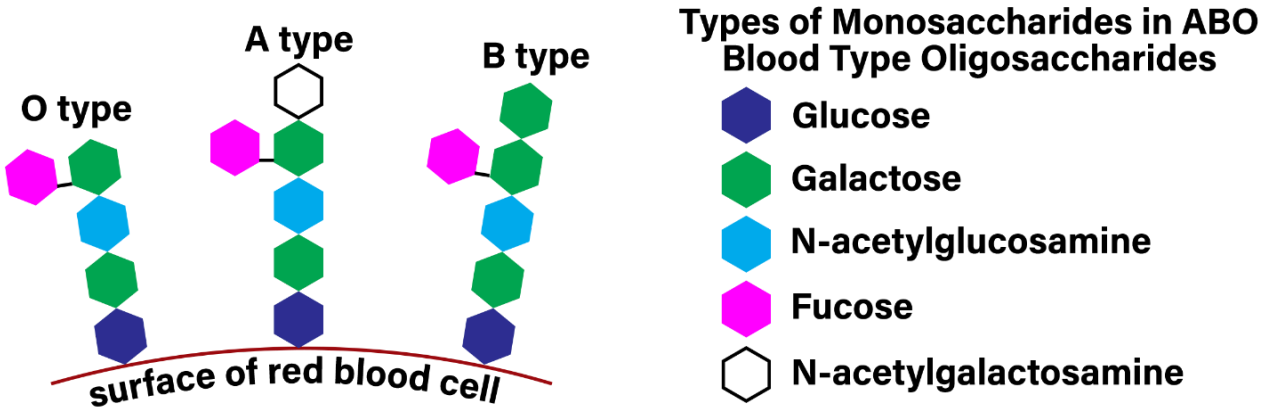
This image above shows different degrees of branching in amylose, amylopectin, and glycogen.

As mentioned above, some polysaccharides are built for structural support. Cellulose is an important structural molecule in plants. Cellulose is a polymer of glucose but is assembled using different glycosidic linkages Humans do not have the enzymes to hydrolyze this linkage and cannot digest cellulose. However, as an insoluble fiber, cellulose is still an important part of the human diet as it promotes intestinal health and helps to lower cholesterol (via bile salt removal). Cellulose is especially plentiful in leafy vegetables and in whole grains.

3.1.4

Oligosaccharides

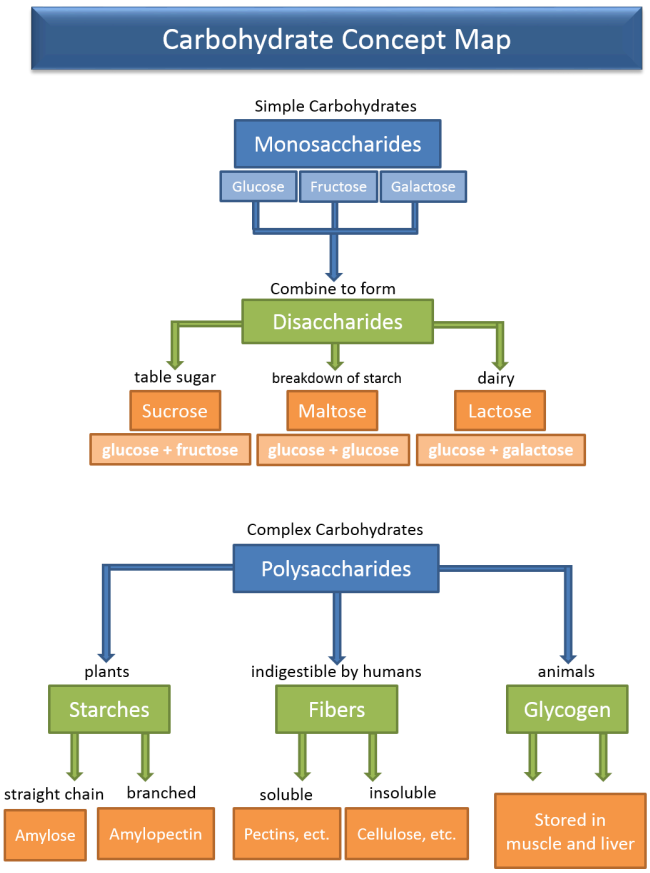
Oligosaccharides differ from polysaccharides primarily in their function. Oligosaccharides range in length from a few monosaccharides up to 100 or more. In humans, most oligosaccharides are between 5-30 monosaccharides in length. These sugars can be linear or branched and are made of a diverse pool of monosaccharides. The diversity in shape and composition facilitate their primary role as cell signaling molecules. Attached to the outer cell membrane, oligosaccharides aid in cell identification and function. A familiar example is blood type. A, B, and O blood types are all designated by a type of oligosaccharide found on the surface of red blood cells (see image below).



ABO blood types are identifiable by different oligosaccharides on the surface of red blood cells.
Image by BYU-Idaho professor Spring 2021

This image shows the different oligosaccharides expressed on the surface of red blood cells that contribute to blood type. These oligosaccharides are branched and are made of 5 different monosaccharides.

In summary, the Dietary Carbohydrate Concept Map shown below ties together the major relationship between and most common examples of monosaccharides, disaccharides and polysaccharides.



Carbohydrate Concept Map: Image created by BYU-I student Hannah Crowder, 2013

Health Note

It is safe to say that carbohydrates are an important part of a healthy diet. Although, some carbohydrates are better than others. When we consume simple sugars, they are quickly absorbed, and blood sugar levels rise rapidly. This, in turn, results in secretion of large amounts of insulin, followed by a rapid drop in blood sugar. This is probably not ideal. Indeed, a recent study1 reported that consuming just one sugary soft drink per day increased the risk of developing coronary heart disease by 20% in men. Consumption of sugar-laden soft drinks has also been shown to increase the incidence of obesity, which increases the risk of type 2 diabetes. Complex carbohydrates found in whole grains, on the other hand, tend to offer positive health benefits.

One current topic of intense interest is the question of high-fructose corn syrup. High-fructose corn syrup is produced from corn starch, which is a polymer of glucose. The starch is hydrolyzed to separate the glucose monomers and then chemically treated to convert some of the glucose to fructose. Most high-fructose corn syrup is 55% fructose and 45% glucose. Fructose is handled by the body differently than glucose. Whereas glucose can enter nearly all cells of the body (some cells need a little help from insulin to take up glucose), fructose is metabolized almost exclusively by the liver. There seems to be mounting evidence that high amounts of fructose may act as a molecular fat switch. As an example, consider a bear who eats lots of berries in the Fall in order to store fat for the impending Winter hibernation. In a recent study, an experiment was done with rats to compare high-fructose corn syrup with sucrose. Rats consuming high-fructose corn syrup had greater weight gain, increased amounts of visceral fat (the fat around our abdominal organs), and an increase in the levels of circulating triglycerides2 (triglycerides are the main component of the fat in our adipose cells). Although there are those that still argue that high-fructose corn syrup is no worse for you than sucrose, the growing body of evidence seems to suggest differently. Therefore, next time you sit down with a nice, cold glass of Sprite, think about what you might be doing to your body.

References

1. Koning, L. de, et al. Sweetened Beverage Consumption, Incident of Coronary Heart Disease and Biomarkers of Risk in Men. Circulation (online). Mar 12, 2012

3. Bocarsly, M.E. et al. High Fructose Corn Syrup Causes Characteristics of Obesity in Rats: Increased Body Weight, Body Fat, and Triglyceride Levels. Pharmacology, Biochemistry, and Behavior. 97:101-106-2012

3.2

LIPIDS

In the Arctic region of the northern hemisphere is a group of people that make up the Inuit culture. The Inuit people have tragically adopted much of the standard American diet (S.A.D.) that we indulge in (fast food, sugary drinks, processed meals, etc.). As a result, the incidence of diabetes and cardiovascular disease is rising dramatically. Inuit natives who adhere to the traditional diet of marine mammals and fish, with some berries and greens, are not demonstrating the startling trend of metabolic disease.
Some call this the Inuit Paradox. The paradox refers to the fact that traditional Inuit diets consist of large amounts of fat. Whale fat, seal fat, caribou fat, and other small animal fat is regularly consumed as a staple. In fact, the daily fat consumption is nearly two times the recommended daily allowance published by the health and nutrition experts in our government. So, why can the Inuit people exist on double the recommended dose of fat intake and actually decrease their incidence of cardiovascular disease and diabetes? This would seem to be a paradox.

There must be more to the unhealthy American diet than the amount of fat consumed. It appears that the type of fat consumed is at least as important as the quantity. Have you heard people talk about saturated and unsaturated fat or vegetable oil and animal fat? Have you heard anything about omega 3 oils and omega 6 oils? What about trans fats and cholesterol? It seems that not all fats and oils are the same. Perhaps the types, mixtures, and ratios of fats and oils found in the traditional Inuit diet can help explain the health benefits enjoyed by traditional Inuit natives.

Science is taking a closer look at other traditional diets as well. Have you heard the hype about the heart protective effects of the Mediterranean diet? Some feel that the types of oils and fats found in this part of the world may have heart protective effects. You will likely hear and read more on the topics of fats and oils in years to come. In order for you to be a literate consumer of this information, it will be important that you know your fats. This section will teach about the different types of lipids.

The Nature of Lipids

As lipids include a vast array of naturally occurring organic molecules. Lipids can be categorized as fats, oils, waxes, cholesterol, cell membranes, some pigments, some vitamins, and many other important compounds. Among the many types of lipids, the terms "fat" and "oil" are probably the most familiar. Fats are generally solid at room temperature while oils are liquids. Here we will examine 3 primary classes of lipids: triglycerides, phospholipids, and steroid lipids.

Triglycerides
Phospholipids
Steroids
Lipoproteins

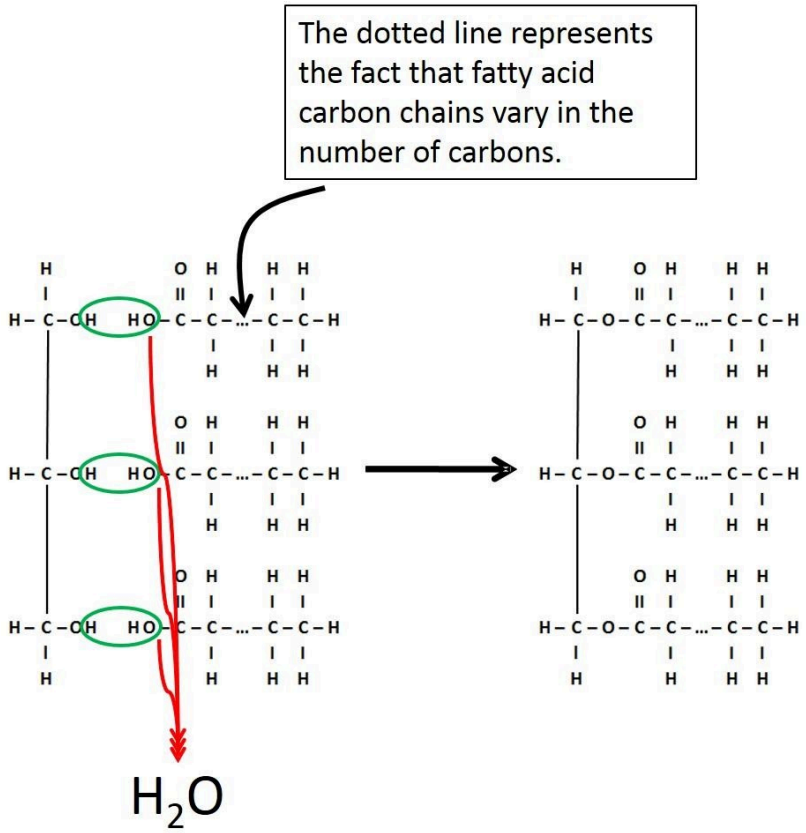
Lipid Profile Values

3.2.1

Triglycerides

Triglycerides (also called triacylglycerol) constitute the major form of fat stored in plants and animals. Over 90% of our dietary fat exists as triglycerides, and over 90% of the fat in a human body resides as triglycerides in adipose tissue below our skin. It is this stored fat that helps cushion and insulate the body. It is also this stored fat that can become excessive and contribute to the many problems of obesity.

Triglycerides are composed of two molecular building blocks: glycerol and fatty acids. Glycerol is a 3-carbon sugar alcohol. A triglyceride is formed by attaching a fatty acid to the hydroxyl group (-OH) of each of these 3 carbons of glycerol through a process called esterification. This reaction is a dehydration synthesis reaction (water is removed) and the resulting bond is called an ester linkage (see figure below).

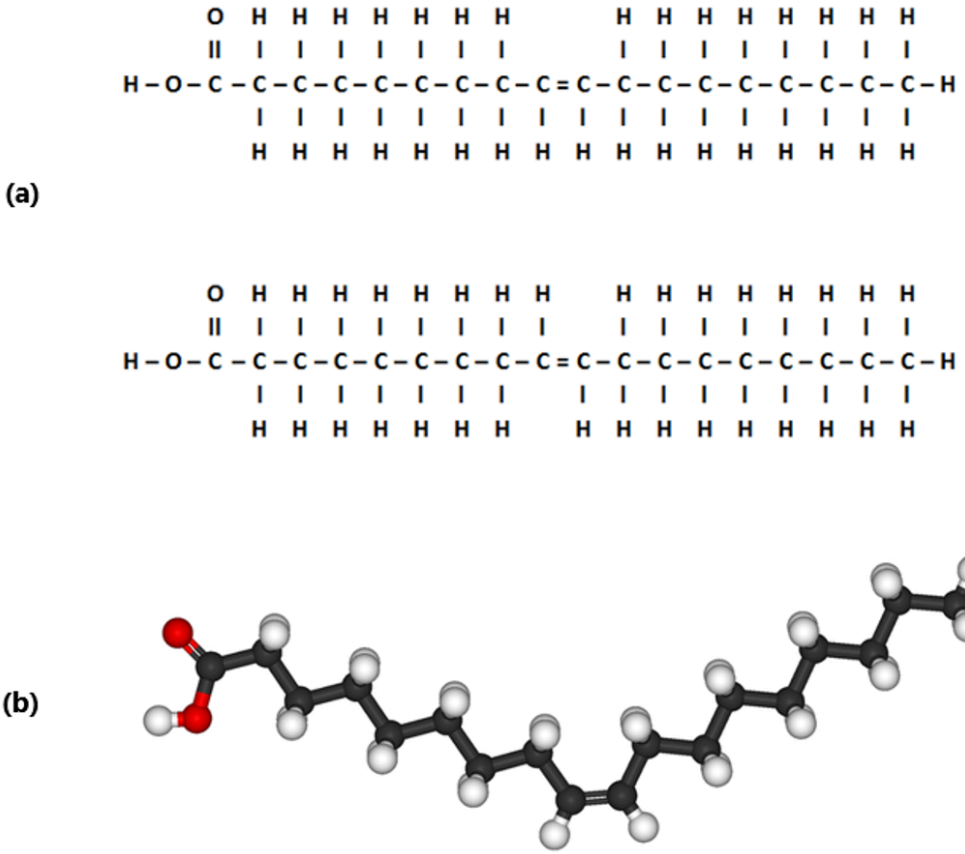


Bonding of Three Glycerol and Fatty Acids by Dehydration Synthesis Reaction to Form Triglyceride.
Image created by JS at BYU-Idaho 2014

To understand the properties of a triglyceride, you must first understand the properties of fatty acids. Fatty acids are hydrocarbon chains with a carboxyl group (-COOH) at one end. The hydrocarbon chain consists of carbon-carbon and

carbon-hydrogen bonds, which are non-polar covalent bonds and therefore hydrophobic. The carboxyl group at the beginning of the hydrocarbon chain is considered a weak acid because it can donate a proton at physiologic pH. Hence the name “fatty acid”.

Fatty acid chains can vary in length, as well as the number and type of carbon-carbon double bonds contained within the hydrocarbon chain. Fatty acid chains with no carbon-carbon double bonds are referred to as saturated. This means that every carbon-carbon bond in the chain is a single bond, which allows two hydrogen atoms to link to every carbon in the chain, except for the last carbon which is bonded to three hydrogen atoms. However, if a double bond occurs between two carbons in the hydrocarbon chain, then the carbon atoms connected by a double bond will each bond with one less hydrogen atom in order to maintain four bonds per carbon atom. As such, the hydrocarbon chain is no longer "saturated" with hydrogen atoms at every carbon. Therefore, an unsaturated fatty acid will contain one or more double bonds (see the image below).

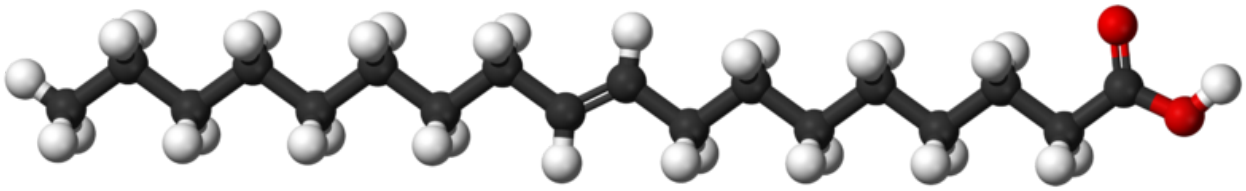


(a) Cis & Trans Double Bond in Monounsaturated Fatty Acid; (b) Cis Double Bond in Unsaturated Fatty Acid.
Image created by JS at BYU-Idaho 2014: Modified File: Oleic-acid-3D-ball-&-stick.png; Author: Benjah-bmm27; Site: https://commons.wikimedia.org/wiki/File:Oleic-acid-3D-ball-%26-stick.png; License: Public Domain

Carbon-carbon double bonds significantly affect the behavior of the fatty acid. These double bonds can occur in one of two states: cis (same) or trans (across). The figure above shows a line drawing of two monounsaturated fatty acids. Note that the first molecule in illustration (a) has the hydrogen atoms extending from the carbon chain on the same side of the double bond. This is called a cis double bond . Note that the second molecule in illustration (a) is nearly identical except that the hydrogen atoms extend from the carbons at the double bond on opposite sides. This is called a trans double bond. Illustration (b) shows a 3D representation of the cis double bond in the unsaturated fatty acid chain. Notice how cis bonds bend or put a kink in the carbon chain.

A fatty acid with one double bond is referred to as a monounsaturated fat, and fatty acids with two or more double bonds are polyunsaturated fats. All lipid containing foods have a specific mixture of saturated and unsaturated fatty

acids. Because saturated fatty acids tend to be straight, they can pack together more tightly. The more tightly packed molecules of fat are more dense and more likely to be solid at room temperature. In contrast, unsaturated fats with cis bonds allow for a kinked or angled geometry that makes it more difficult to pack together, causing them to be a liquid at room temperature. Most naturally occurring unsaturated fats are cis fats. Commercial trans fats, which retain a straight shape and behave like saturated fats, have been banned from use in culinary production due to an increased risk of heart disease associated with long-term consumption.



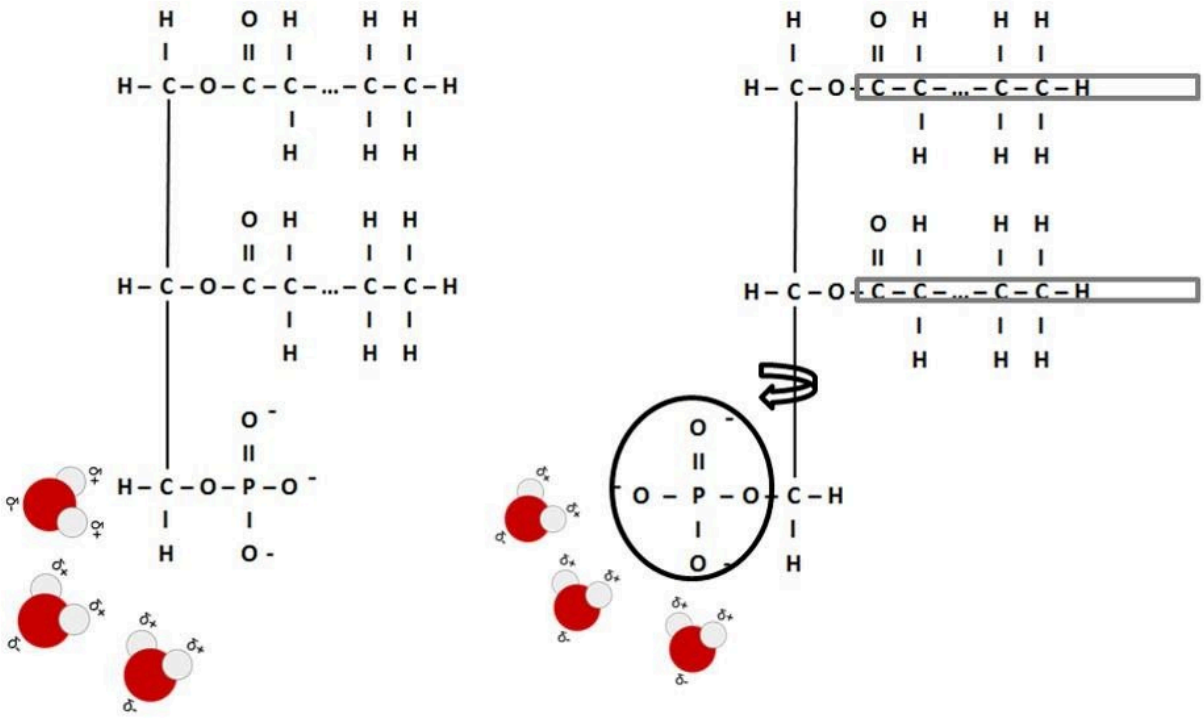
Trans Double Bonds in Unsaturated Fatty Acid. Title: File: Tridecylic-acid-3D-balls.png; Author: Jynto and Ben Mills; Site: https://commons.wikimedia.org/wiki/File:Tridecylic-acid-3D-balls.png; License: public domain

Epidemiologic, clinical, and physiological studies are quite consistent in showing that saturated and trans fats cause adverse effects on our health if they are eaten in excess. Throughout the world, trans fats have been acknowledged as unhealthy, and many countries have outlawed their use. Substituting unsaturated fat for saturated and trans fat has sometimes been shown to lower LDL cholesterol which may have a "heart healthy" effect. There is also mounting evidence that some unsaturated fats may have the ability to decrease chronic inflammatory responses in the body. Omega 3 fatty acids have come into the spotlight lately as they may help to do this. Omega 3 refers to the existence of a double bond three carbons in from the end of a fatty acid chain.

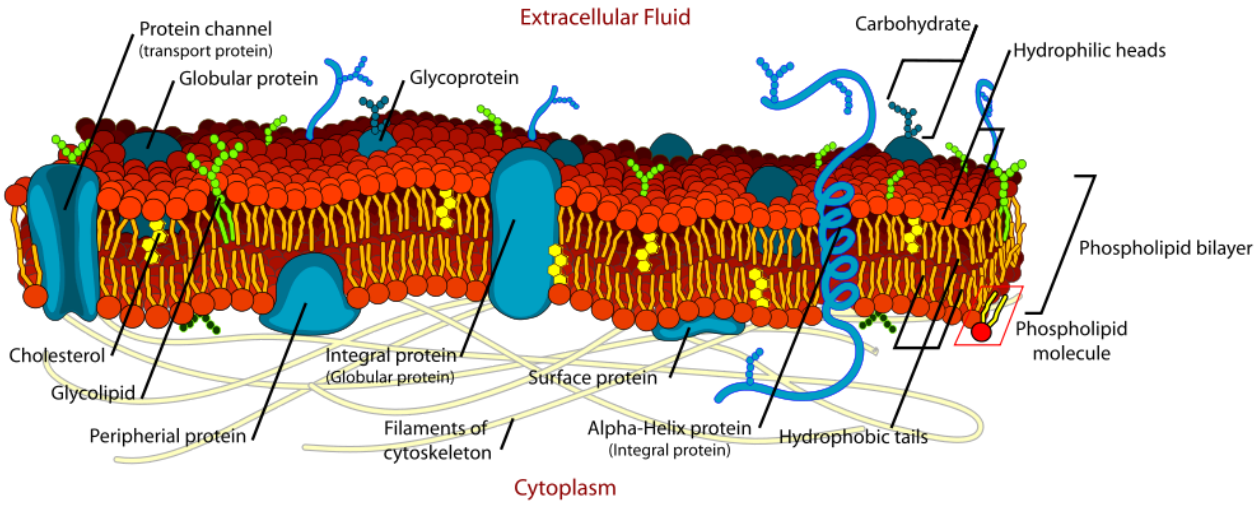
3.2.2

Phospholipids

A phospholipid is structurally similar to a triglyceride. However, one of the three fatty acids is substituted for a polar phosphate group (see image below). This creates a molecule that is hydrophilic on one end (the phosphate group) and hydrophobic on the other end (2 fatty acid tails). This type of molecule – hydrophobic on one end and hydrophilic on the other – is referred to as amphipathic. The amphipathic nature of phospholipids allows them to maintain the structural integrity of a cell membrane and serves as a selectively permeable barrier that modulates movement of substances in and out of a cell.



Phospholipid Structure Showing Polar Phosphate Group. Image created by JS at BYU-Idaho 2014; modified File: Na+H2O.svg; Author: Taxman; Site: https://commons.wikimedia.org/wiki/File:Na%2BH2O.svg; License: Public Domain
Because of their amphipathic nature, phospholipids spontaneously coalesce into spheres (called micelles) when placed in water. In like manner, a double layer of phospholipids, called a lipid bi-lipid layer, constitutes a cell membrane (see figure below).

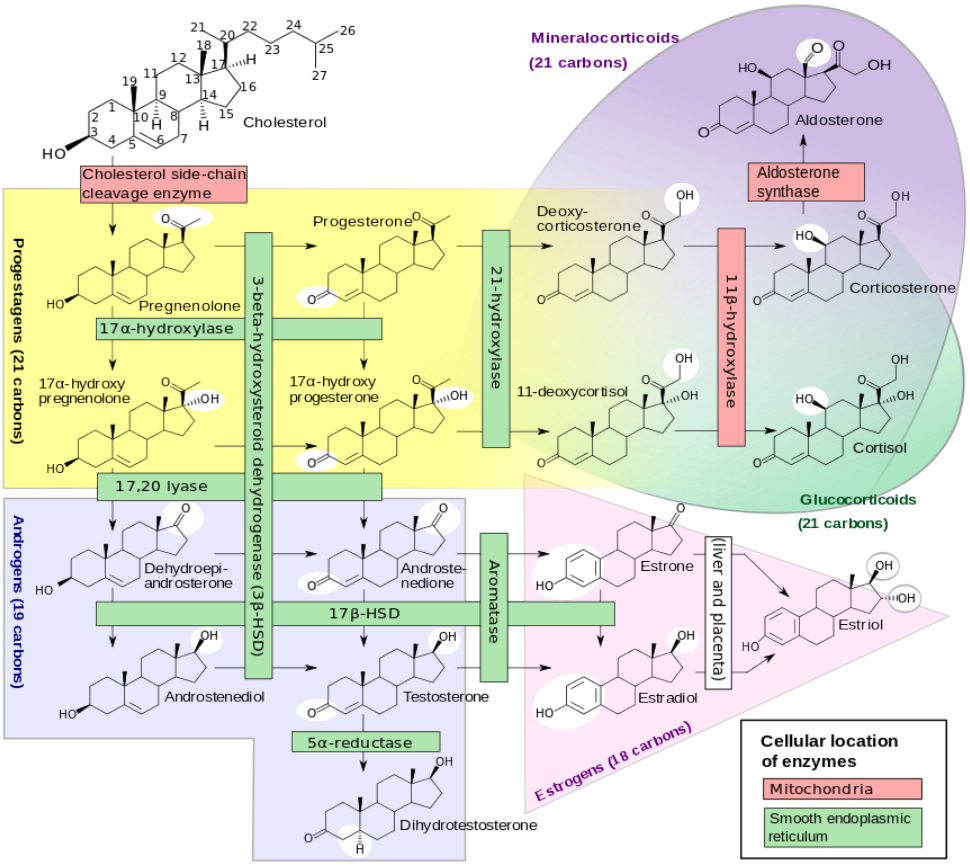


Cell Membrane Detailed Diagram File: Cell membrane detailed diagram en.svg; Author: LadyofHats Mariana Ruiz; Site: https://commons.wikimedia.org/wiki/File:Cell_membrane_detailed_diagram_en.svg; License: Public Domain
In a cell membrane the polar heads of the phospholipids are oriented towards the aqueous cytoplasm and also towards the extracellular water. The fatty acid tails are oriented away from water but blend with each other. This configuration creates a barrier or boundary that separates the cytoplasm environment from the extracellular environment. Cell membranes also contain proteins and cholesterol (a steroid lipid) which aid in attachment and signaling, and also membrane integrity (see image above).

3.2.3

Steroids

A **steroid lipid** is a type of lipid that differs in structure from triglycerides and phospholipids. A steroid lipid consists of 4 hydrocarbon rings (3 hexamers and a pentamer) that are joined to each other (see image below). However, the solubility characteristics of steroids are like other lipids in that they are nonpolar (hydrophobic). The best known and most abundant steroid lipid in the body is **cholesterol**. Cholesterol is very important for several reasons. First, it is required to build and maintain cellular membranes. Second, cholesterol is used to synthesize bile, an important component of digestive juices that helps in the digestion of fat. Third, cholesterol is also used to synthesize steroid hormones (see image below). Steroid hormones are critical for healthy growth and development of most tissues in our body. Cholesterol is essential to all animal life, so we find that animals (including humans) have the ability to make this important molecule. We can also ingest cholesterol. When we consume animal products, we obtain cholesterol to varying degrees. Cheese, egg yolks, beef, and pork are all examples of foods commonly considered to have substantial amounts of cholesterol. Excess cholesterol may contribute to the formation of deposits on the inner walls of blood vessels. These deposits can become quite hard and can result in a condition known as atherosclerosis.



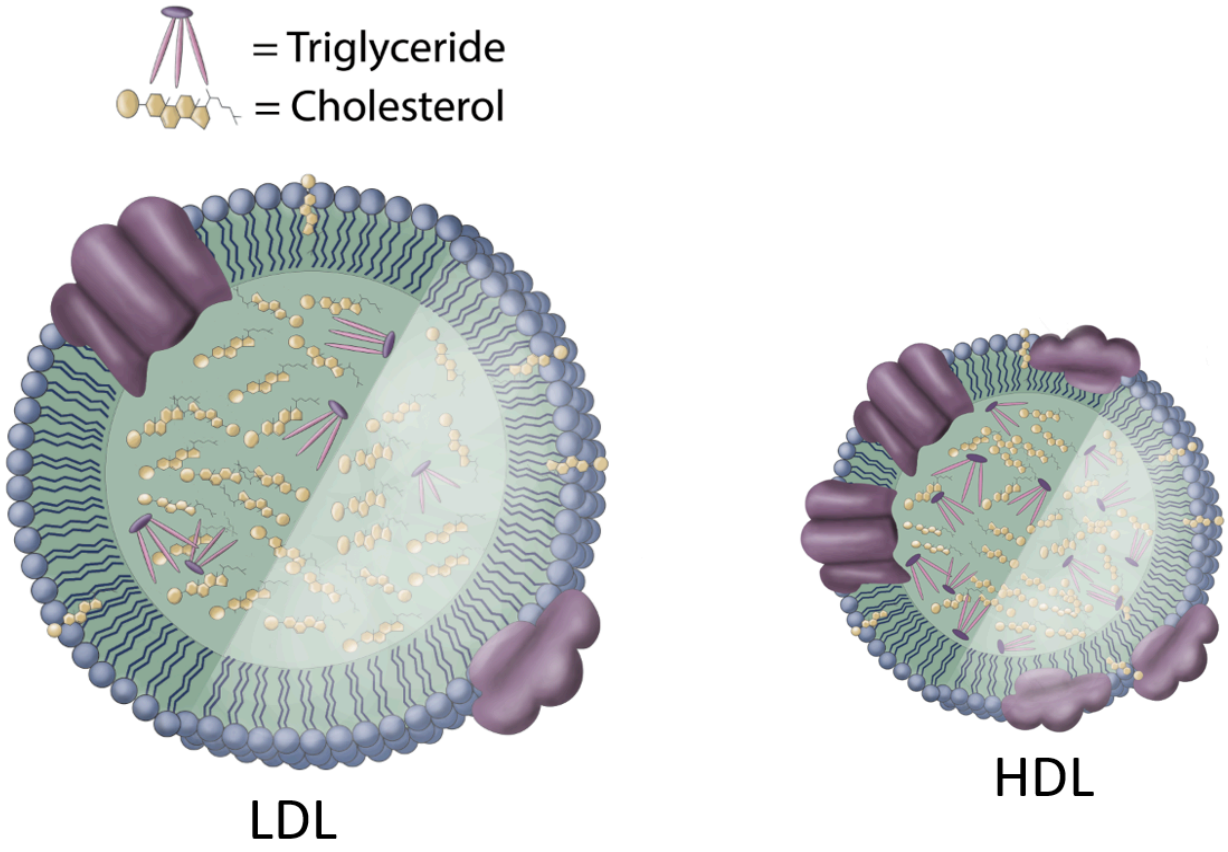
Steroideogenesis. *File: Steroidogenesis.svg; Author: Hoffmeier and Setters. Site: https://en.wikipedia.org/wiki/File:Steroidogenesis.svg; License: Creative Commons Attribution-Share Alike 3.0 Unported License.*

This figure shows man of the steroid hormones that are synthesized from cholesterol, including estrogen, progesterone, testosterone, and cortisol.

3.2.4

Lipoproteins

The transport of lipids (triglycerides and cholesterol) throughout the body is important to provide energy and building blocks to cells. However, because lipids are hydrophobic, these molecules do not easily dissolve in blood. To transport lipids through the blood, the body uses micelle-like structures called **lipoproteins**. As discussed in the previous section on phospholipids, a micelle is a sphere consisting of a single layer of phospholipids. A lipoprotein is essentially a micelle with certain proteins embedded within the phospholipid monolayer. As the phospholipid hydrophobic tails orient toward the inside of the sphere, this hollow structure becomes a useful tool as it transports lipids within its hydrophobic lumen. Cholesterol and triglycerides travel inside of these spheres and are shielded from the water (see figure below). This lipoprotein particle will dissolve in the water portion of blood plasma and be carried easily through the circulation. Lipids will not form droplets or plug up any vessels because they are safely tucked away in the lipoprotein assembly. Cells throughout the body have the ability to bind these lipoproteins and move lipids in and out of them. There are several different types of lipoproteins, and each one has a specific role. Some will focus on picking up lipids from the digestive tract, while others will specialize in the transport of lipids between the liver and body tissues.



Low-Density Lipoprotein and High-Density Lipoprotein. *Image created by Dessa Selch BYU-Idaho, 2013.*

Above is a schematic representation of two types of lipoproteins, with the phospholipid molecules shown in blue and the apoprotein molecules shown in orange. LDL stands for Low-Density Lipoprotein and HDL is an acronym for High-Density Lipoprotein. As has been mentioned above, lipoproteins are an assembly of phospholipids and proteins that carry triglycerides and cholesterol in the blood. Low-Density Lipoprotein (LDL) is lower in protein and higher in lipid content, and High-Density Lipoprotein (HDL) has more apoproteins which increases this particle's overall density. The more protein that a lipoprotein assembly contains, the heavier it is (see image above).

You may have heard of these terms **HDL** and **LDL** before. These are two common lipoproteins that have gained a lot of attention as they appear to correlate with the risk of atherosclerosis development. Many brochures and websites refer to HDL as "Good Cholesterol" and LDL as "Bad Cholesterol."

In reality, there is no such thing as "good cholesterol" or "bad cholesterol." Cholesterol is simply a type of lipid that is necessary for life. It does not come as "bad" or "good." The idea of "good" and "bad" refer to the lipoproteins. LDL particles tend to accumulate in the walls of arteries. It is the overabundance of this LDL deposition that contributes to atherosclerosis, hence why it receives the term "bad." HDL or High-Density Lipoprotein is often called the "good cholesterol" because HDL particles help prevent atherosclerosis by extracting cholesterol from artery walls and disposing of it through biochemical reactions in the liver. Research has shown that lowering LDL cholesterol reduces the risk of heart attacks, strokes, and atherosclerosis.

LIPID SUMMARY

The lipids include a large and diverse group of naturally occurring compounds. Triglycerides, phospholipids and cholesterol are three important lipids in biology. The large stores of fat in our bodies are mostly triglycerides, which also comprise the bulk of fats and oils that we consume. Triglycerides contain saturated and unsaturated fatty acids that lend physical and chemical properties such as solidity, texture and flavor to food. Fatty acids also affect health. Saturated fatty acids appear to contribute to heart and vascular disease when consumed in high quantities. Unsaturated fatty acids may actually lower some health risks like inflammation and heart disease.

3.2.5

Lipid Profile Values

The following appendix offers information regarding desirable and undesirable lipid profile values.

Cholesterol Lipid Panel Results: *Image Permission Acquired by email through Nate Wise BYU-Idaho Dec 2013*

Total Cholesterol (TC)	**Category**
Less than 200	Desirable
200-239	Mildly High
240 and above	High

HDL	HDL- Category
60 and above	High; Optimal; associated with lower risk
Less than 40 in men and less than 50 in women	Low; considered a risk factor for heart disease

Triglycerides	Triglyceride Category
Less than 150	Normal
150-199	Mildly High
200-499	High
500 or higher	Very High

LDL	LDL Category
Less than 100	Optimal
100-129	Slightly Above Optimal
130-159	Borderline High
160-189	High
190 and above	Very High

Non-HDL	Non-HDL Category
Non-HDL is a reading that includes the cholesterol content of all the lipoproteins that are not part of the HDL classification. LDLs are the most common lipoprotein to examine for heart disease risk, but there are other lipoproteins that can contribute to atherosclerosis. These are sometimes called Very *Low-Density Lipoproteins* (VLDL) and *Intermediate Density Lipoprotein* (IDL). The general category of all Non-HDL lipoproteins can be combined with the readings from the LDL category to validate concerns for heart risk profiles.	

Less than 130	Optimal
129-159	Slightly above Optimal
160-190	High
Above 190	Very High

TC/ HDL	**Category**
Total Cholesterol to HDL ratio (TC /HDL) is a number that reflects how many HDL lipoproteins we have relative to our total cholesterol. A person with a lower HDL value may see that his/her total cholesterol is also low. In this case, a person with a lower HDL value may have a TC/HDL ratio that is fine. This value taken with the other values in the lipid profile help a health care professional get a better idea of the actual heart disease risk.	
Below 3.5	Optimal
3.6 to 5.0	Borderline High to High
Above 5.0	High to Very High

12-hour Fasting Glucose (Glu)	**Category**
82 - 110	Optimal
111- 125	Borderline High to High
Above 126	High to Diabetic

3.3

PROTEINS

Of the four classes of biological molecules, the proteins are the most diverse in their functions. By some estimates, our cells make more than 50,000 different proteins, with each protein having a specific job within the body. Consider lactase, whose job it is to break down lactose into glucose and galactose within the small intestine. Lactase binds specifically to lactose. It won't break down sucrose or maltose. Just lactose. And if the lactase enzyme is absent or broken? Talk to someone who is lactose-intolerant about their symptoms.

With such diversity, what gives a protein its functionality and specificity? For proteins, form is function. In other words, the specific 3-dimensional shape of a protein is what allows it to do its job. Table 1 lists some of the major functions of proteins, but this list is not exhaustive. In fact, it is hard to think of any function in the body in which proteins are not integral. In this unit, we will learn about the molecular structure of proteins and discuss some of their important functions.

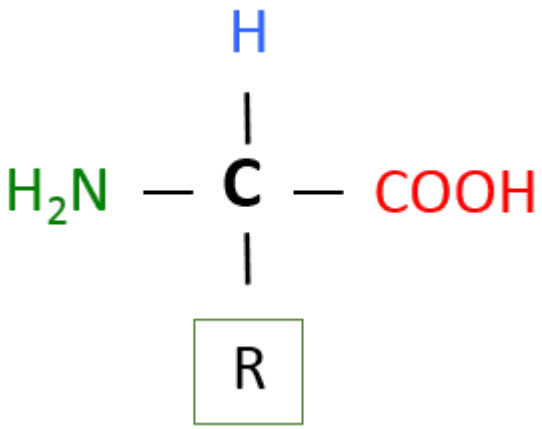
Function	**Example**
Structure	Collagen in tendons and ligaments, Keratin in the nails and skin
Transport	Hemoglobin in the blood, Na^+, K^+-ATPase in cell membranes
Protection	Antibodies of the immune system
Movement	Actin and Myosin in muscles
Enzymes	Digestive enzymes in the small intestine (Lactase, Sucrase, Trypsin)
Receptors	Membrane proteins that respond to chemical messengers (insulin receptors)
Regulation	Chemical messengers: hormones, neurotransmitters, cytokines

Amino Acids
Peptide Bonds and Polypeptides
Protein Structure
Classes of Proteins
Enzymes

3.3.1

Amino Acids

Like polysaccharides and nucleic acids, proteins are polymers of smaller subunits or monomers. The monomers that make up proteins are the **amino acids**. Although there are 20 different amino acids that make up the proteins in humans, all have the same basic structure. Each has a central carbon with four different groups attached to it. The figure below shows the basic structure of an amino acid. Attached to the central carbon is an **amine group** (-NH2) and a **carboxyl group** (-COOH). The carboxyl group acts as a weak acid (a proton donor) at physiologic pH. The name amino acid is derived from these two groups. Additionally, there is an **R group**. In organic chemists' shorthand, the R group represents some other organic group. In the case of the amino acids, there are 20 different R groups, hence 20 different amino acids. It is the different R groups that confer the different properties to the amino acids. Some R groups are nonpolar (hydrophobic); others are polar (hydrophilic). Some R groups contain ions, either anions or cations (hydrophilic). In this image, the R group simply represents some type of organic group. Understanding the properties of R groups is important to understand protein folding and 3-dimensional shape.



Amino Acid Structure: Carbon (C), Amine Group (-NH2), Carboxyl Group (-COOH), Hydrogen (-H), and One of 20 various Amino Acid R Groups. *Image created by MG BYU-Idaho 2013*

In this class, you will not need to learn the names and structures of the individual amino acids. However, if you are interested in learning more about their structures and characteristics, check out the following links:

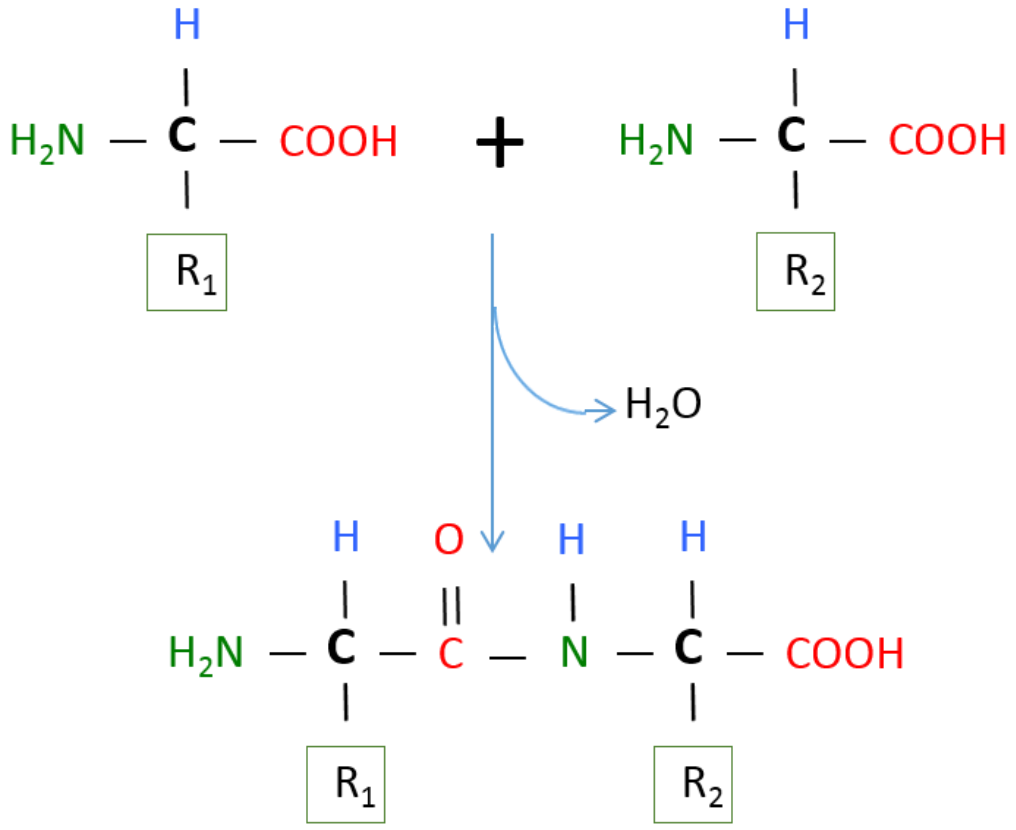
http://www.aminoacidsguide.com/

https://books.byui.edu/-sMw

3.3.2

Peptide Bonds and Polypeptides

As mentioned in the introduction, proteins are polymers of amino acids. Like all of the polymers we have discussed so far, amino acids are linked together via **dehydration (condensation) synthesis reactions**. The bond that is formed between the amino acids is called a **peptide bond**. The figure below shows how these bonds are formed. In this simple example, we would call the resultant polymer a **dipeptide**. Small peptides are designated tripeptides, tetrapeptides, pentapeptides, etc. The generic term polypeptide is used to designate many amino acids linked together. The terms polypeptide and protein are often used interchangeably. A polypeptide chain has at its beginning an unbound amino group and is given the name **amino- or N-terminus**, while the other end of the chain is called **carboxyl- or C-Terminus**.



Peptide Bond formed through Dehydration Synthesis of Amino Acids. *Image created by MG BYU-I; 2013.*

The image above represents a dehydration synthesis reaction between two amino acids to form a peptide bond. Peptide bonds form between the carboxyl group of one amino acid and the amine group of another.

As mentioned above, almost all living things contain proteins made from 20 amino acids. Our liver is a pretty effective amino acid factory and can synthesize 11 of these 20 amino acids even if we don't consume them in our diet. However,

nine amino acids are **essential amino acids**. If we don't consume these essential amino acids in our food, our bodies won't have the necessary supplies available when new proteins need to be produced. These essential amino acids are histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine.

To review, think of amino acids as Lego blocks. If you were given 20 packages of these blocks, each package containing a different color, you could start producing Lego proteins. Just think of the possible combinations of your Legos. The possibilities are essentially limitless. Some of your proteins may contain only a few Legos, while others may contain thousands, each with a different shape and color combination. This is the potential that our cells have at their fingertips to produce the molecules to carry out the many functions of proteins.

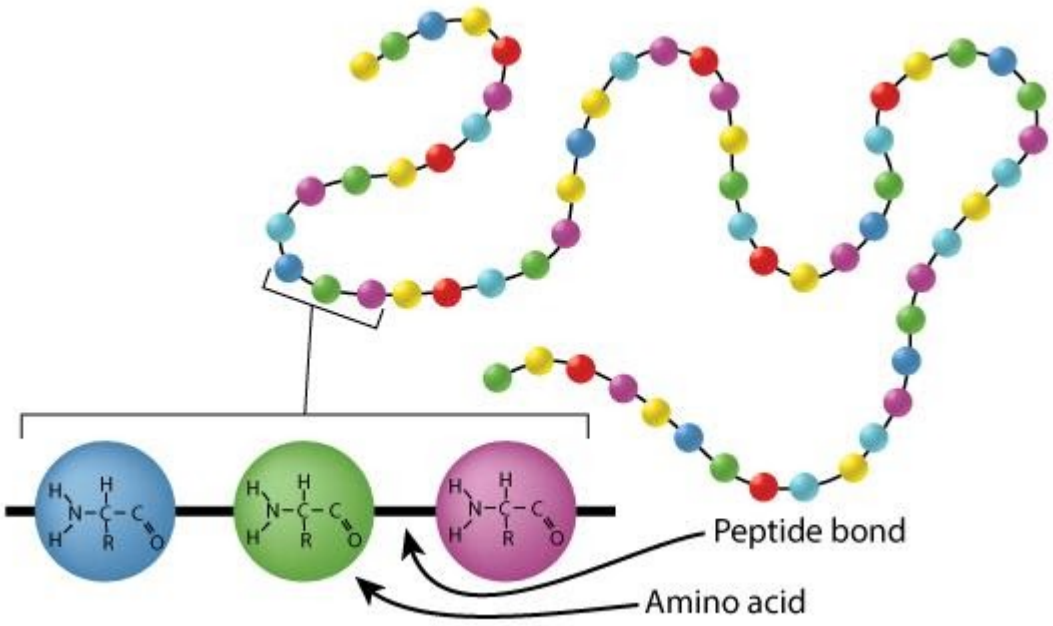
3.3.3

Protein Structure

By now, you should be starting to realize the importance of proteins; they are critical to the proper functioning of various bodily systems. What is it about proteins that allow them to perform all of these different tasks? The answer to this question can be summed up in three words: shape, shape, and shape. As mentioned earlier, form (shape) equals function. As you can imagine from the many functions of proteins, they have very complex shapes. If we think of proteins as cars, we all quickly understand that the wheels on the bottom of the car and a steering wheel to guide the car are very important standard equipment. Similarly, if our protein doesn't have the right parts in the right places with each component properly connected together, the protein does our body about as much good as a car that has been put through an auto crusher. In studying the shape of proteins, biochemists have dissected and broken them down into four levels of complexity or structure. Keeping with the car analogy, if we really wanted to dissect a car and determine how it works, we could take it apart all the way down to the nuts and bolts and then reassemble it again. Biochemists do the same thing to proteins to try and understand how proteins work. The first level would be analogous to the “parts” level. As we move from the first to the fourth level of structure, the preceding level adds to the next. For example, you cannot have secondary structure without a primary structure.

Primary Structure (First Level)

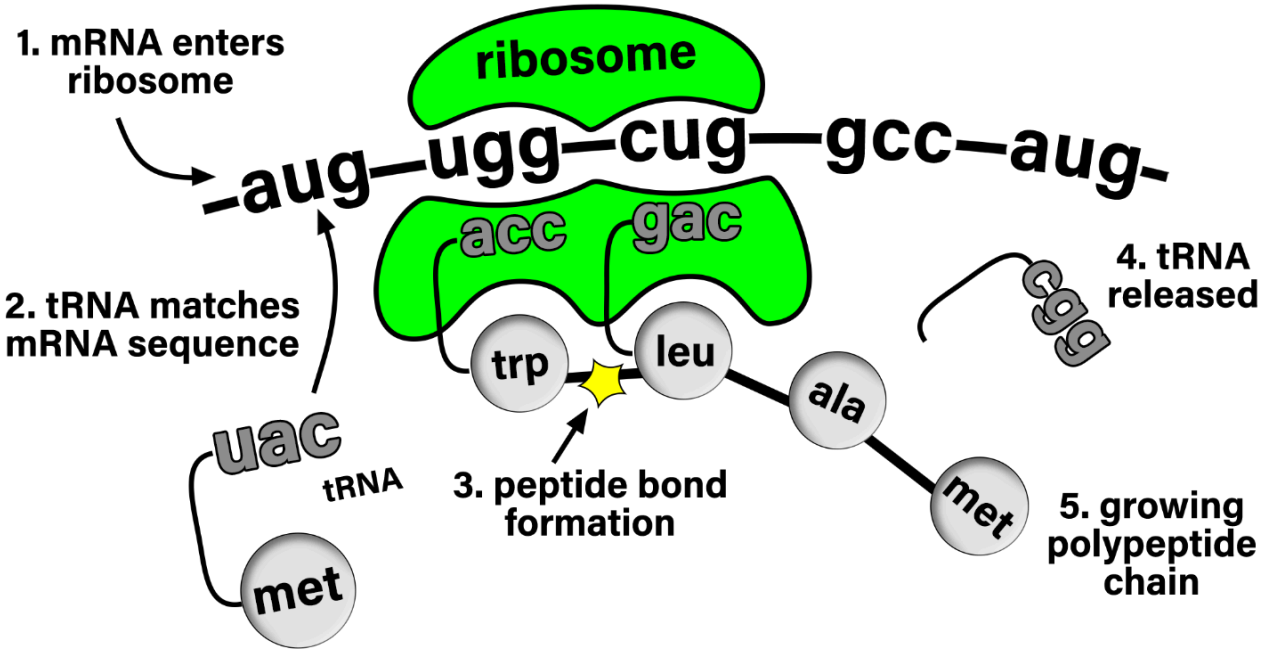
The primary structure of the protein is the sequence of the amino acids in its polypeptide chain. If proteins were popcorn stringers made to decorate a Christmas tree, the primary structure of a protein is the sequence in which various shapes and varieties of popcorn are strung together. The primary structure of a protein is maintained by **covalent peptide bonds** connecting the amino acids together. Insulin, the first protein to be sequenced, contains the following 110 amino acid primary sequence: malwmrllpl lallalwgpd paaafvnqhl cgshlvealy lvcgergffy tpktrreaed lqvgqvelgg gpgagslqpl alegslqkrg iveqcctsic slyqlenycn. Each letter is specific for 1 of the 20 amino acids.



Primary Protein Structure: Insulin Polypeptide Chain linked by Covalent Peptide Bonds.

Image by BYU-Idaho student Nate Shoemaker Spring 2016

The image above represents the primary structure of a protein (a chain of amino acids). As you might expect, the sequence of the amino acids in the polypeptide chain is crucial for the proper functioning of the protein. Importantly, how does the cell know the right order in which to connect the amino acids? The original code is found in the DNA (deoxyribonucleic acid) housed in the nucleus of the cell. When a specific protein needs to be made, a segment of DNA called a gene is first copied in a process called transcription. This copy is called messenger RNA (mRNA). The mRNA strand exits the nucleus and attaches to a ribosome, a specialized organelle within the cell that interprets the code contained in the mRNA, recruits the appropriate amino acid, and catalyzes the formation of the peptide bond that links amino acids together. The process is called translation and results in a growing polypeptide chain (see figure below).



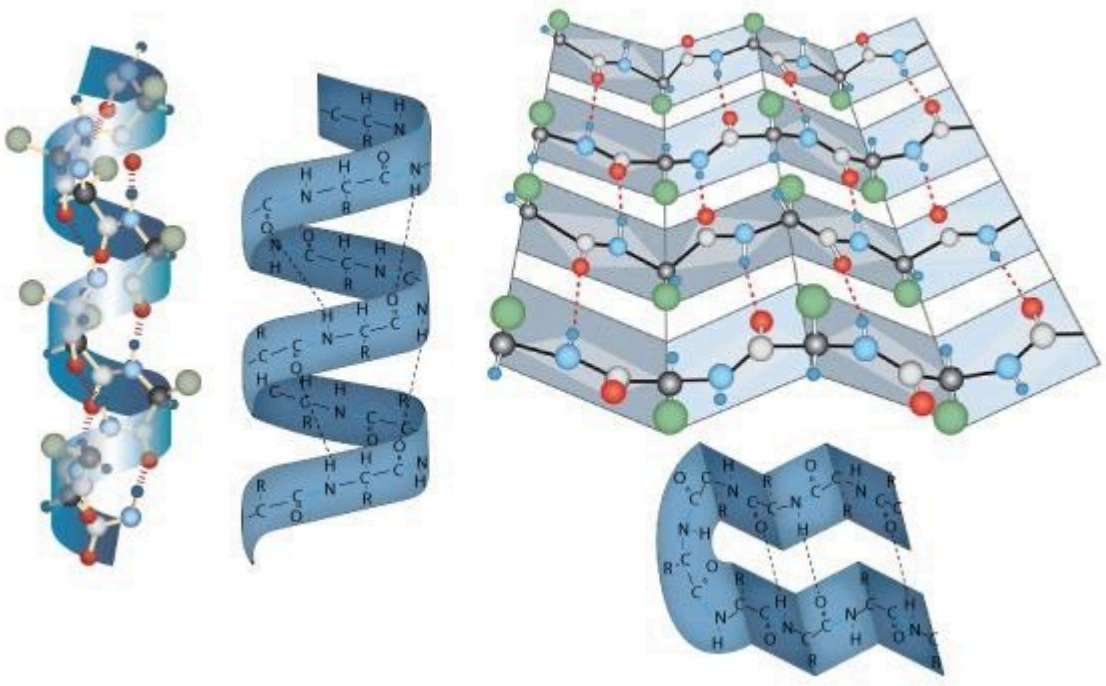
Protein Translation. The mRNA feeds through the ribosome which helps match the appropriate tRNA carrying its respective amino acid. The ribosome then catalyzes the formation of a peptide bond between amino acids to create a polypeptide chain.

Image by BYU-Idaho professor Spring 2021

As you can now see, if there is a mutation in the DNA then the amino acid sequence may be altered. The function of the protein can be affected. All of the known genetic diseases, such as cystic fibrosis, sickle cell anemia, albinism, etc., are due to mutations that result in alterations in the primary structures of proteins, which then, in turn, cause alterations in the other levels of protein folding: secondary, tertiary, and possibly quaternary structure.

Secondary Structure

The *secondary structure* of proteins involves twisting or folding polypeptides into highly regular sub-structures. Whereas the primary structure of a protein is pretty much two-dimensional, the secondary structure of proteins begins the very important three-dimensional configuration of proteins.



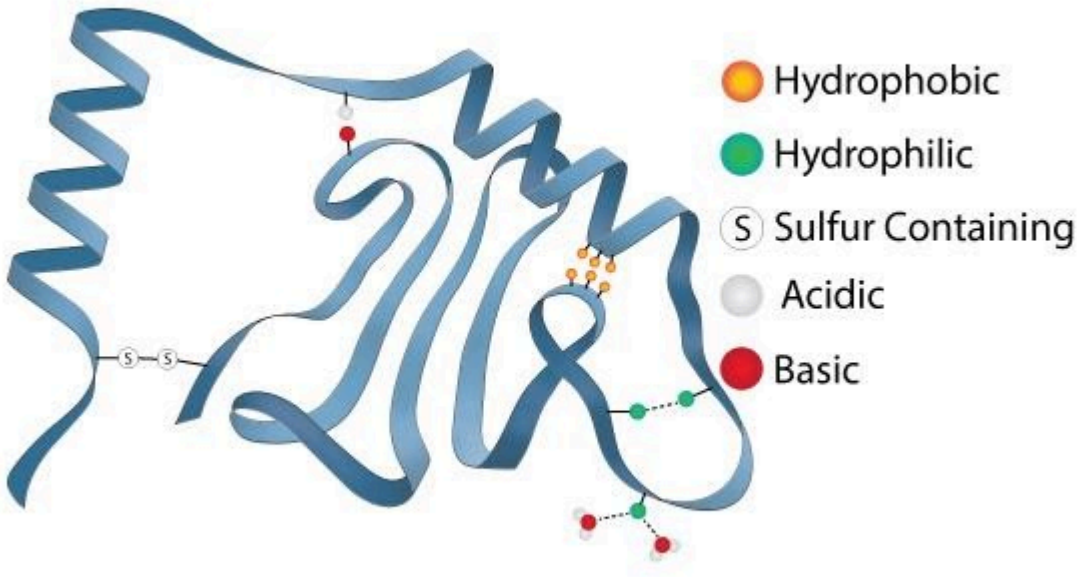
Secondary Protein Structure: Alpha Helix and Beta Pleated Sheet linked by Hydrogen Bonds.

Image drawn by BYU-Idaho student Nate Shoemaker Spring 2016

The two types of secondary structure are the **alpha helix** (think "slinky" as shown in the left picture just above) and the **beta pleated sheet**, or simply pleated sheet (shown to the right in the image above; think about one of those folded cardboard windshield guards that can be placed on the inside of your car's windshield on a hot day so the inside of your car doesn't end up with a temperature approximately that of the interior of our sun). The secondary structure of proteins is a result of the sequence of amino acids in the primary structure and is maintained by **hydrogen bonds**. These hydrogen bods occur along the protein backbone, independent of R-group side chains. Some proteins, like collagen, are almost entirely alpha helix, while others, like silk, are a mostly pleated sheet. Other proteins can have short segments of alpha helix and/or pleated sheet in their structure.

Tertiary Structure

The tertiary structure of a protein is the overall folding of the polypeptide chain, and represents a protein's final 3-dimensional shape. In contrast to secondary structure, tertiary structure can be stabilized by multiple types of bonds (covalent, ionic, hydrogen) and hydrophobic/hydrophilic interactions as dictated by the amino acid R-group side chains.

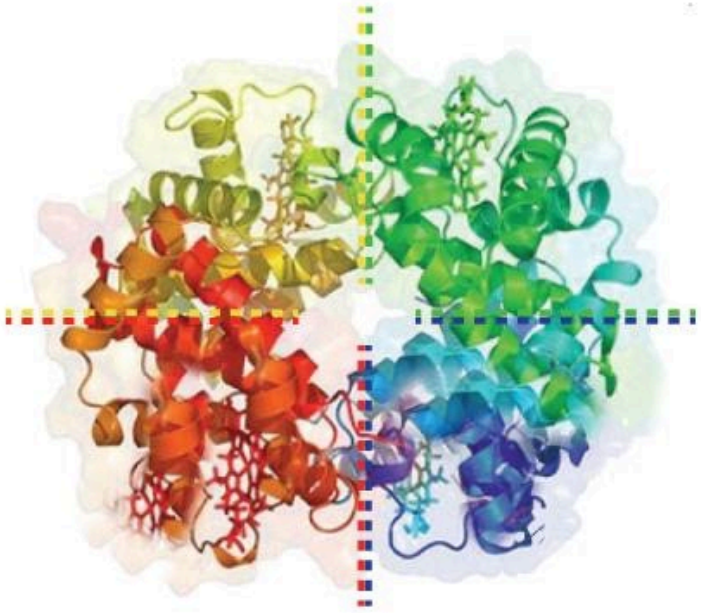


Tertiary Protein Structure: Hydrophilic & Hydrophobic R Groups bound by Hydrogen Bonds, Ionic Bonds Impacted by pH, and Covalent Disulfide Bonds. *Image drawn by BYU-Idaho student Nate Shoemaker 2016*

For example, R-groups that act as weak acids and bases can donate or accept protons. This can create positive and negative charges on the amino acids that will create ionic attraction. Certainly, pH can affect how these attractions between acidic and basic R groups occur. This helps explain why radical changes in pH can cause the structures of proteins to fall apart and ruin the protein's ability to function. One very important and very strong tertiary structure bond is a covalent bond that occurs between R groups on cysteine residues. These R-groups contain sulfur, which can interact with other sulfurs to form a disulfide bridge. The loss of a protein's 3-dimensional shape is called *denaturing the protein.*

Quaternary Structure

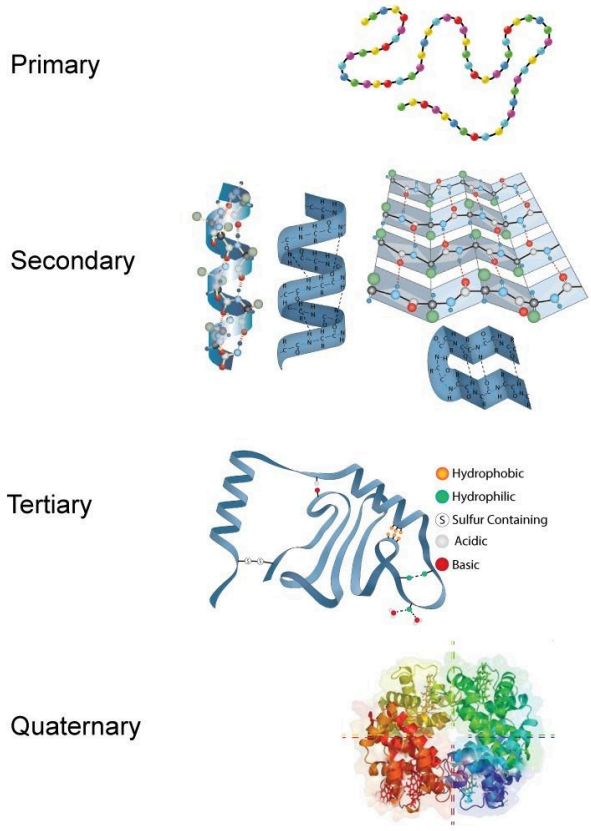
Sometimes multiple protein subunits work together to perform a specific function. Quaternary structure describes the number and arrangement of multiple polypeptide chains coming together to forma functional multi-protein complex. Not all proteins assume a quaternary structure. Only proteins composed of more than one polypeptide chain have quaternary structure. As an example, the protein in the picture below has four polypeptide chains that work together to form one functional protein called hemoglobin.



Quaternary Protein Structure: Four Polypeptide Chains Forming One Protein.
Image developed by BYU-Idaho student Nate Shoemaker Spring 2016

Hemoglobin is found in your red blood cells and has the job of carrying oxygen throughout your body. There are two alpha and two beta chains that make up hemoglobin. You may have heard of sickle cell anemia. This genetic disease is caused by a mutation that results in a change to just one amino acid in the primary structure of the beta chains. This small change is enough to cause a significant alteration to the quaternary structure of hemoglobin, resulting in an abnormal sickle shape. This alteration affects hemoglobin's ability to function correctly, resulting in multiple pathological symptoms.

This next image below is just a summary that shows all the levels of protein structure in one image. You can see how each level leads to a more complex development of a very specific three-dimensional protein.



Four Levels of Protein Structures. *Image developed by BYU-Idaho student Nate Shoemaker Spring 2016*

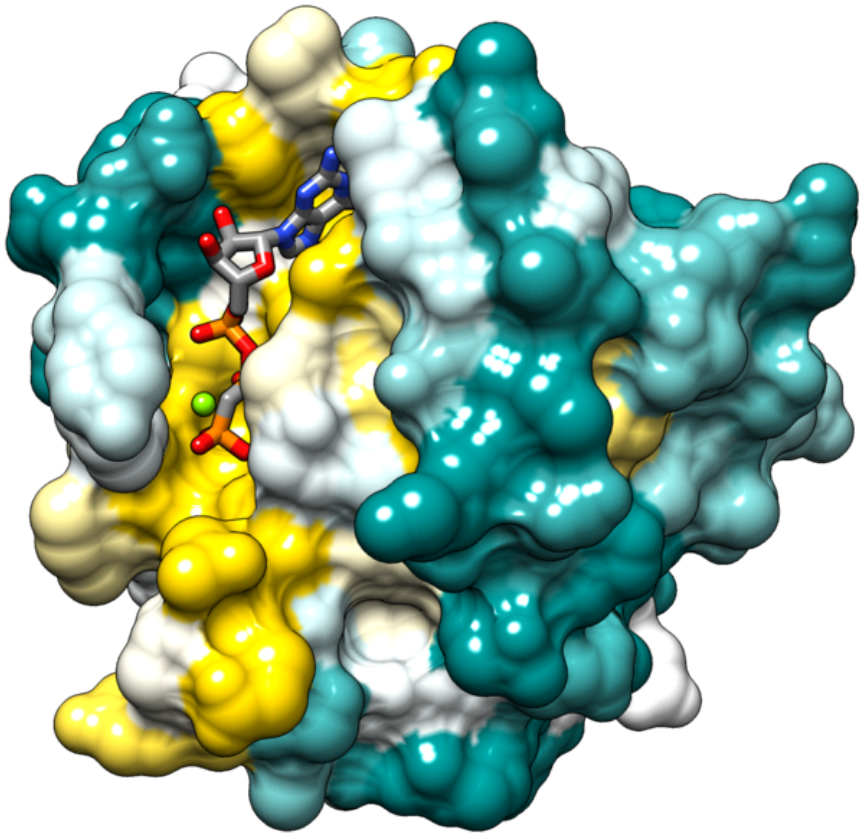
3.3.4

Classes of Proteins

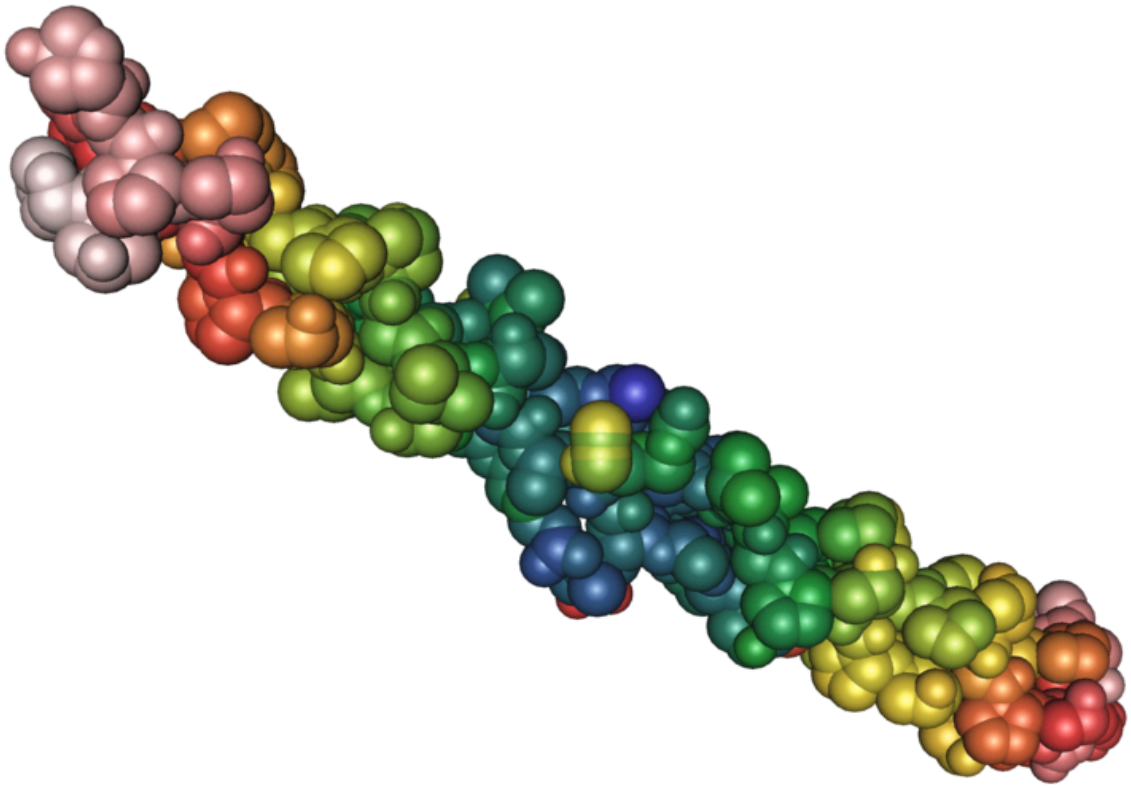
There are two major classes or types of proteins: **globular** and **fibrous proteins**. Globular means globe-like. Hemoglobin is a good example of a globular protein. Globular proteins are quite fragile and can be inactivated (**denatured**) by things like heat (think of the protein albumin in an egg white when you fry it), organic solvents, or strong ionic solutions.

Fibrous proteins are much stronger and tougher. As the name implies, these proteins are more like ropes or cables. Fibrous proteins give the body structural support and help it resist mechanical stress. Common examples of body structures containing fibrous proteins include bone, cartilage, tendons (which anchor muscles to bones), ligaments (which anchor bones to other bones), and capsules around our internal organs.

The two images below show the molecular image representations of a globular (first) and a fibrous protein (second).



Globular Protein. *File: Hras surface colored by conservation.png; Author: Elaine Meng;*
Site: https://books.byui.edu/-nhlb;
License: licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license.



Fibrous Protein. *File:1bkv collagen 02.png; Author: Nevit Dilmen;*
Site: https://books.byui.edu/-RGFQ;
License: licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license.

3.3.5

Enzymes

One of the most diverse and important class of proteins is that of enzymes. An enzyme's purpose is basically to speed up the rate of a chemical reaction. Enzymes accomplish this by decreasing the activation energy needed to start the process. As an example, consider a wooden stick just thin enough to be broken in half if you exert all of your strength. Now, take that stick and place it in a vice or a press and start compressing the stick until it is just about to break. Under this pressure, how much energy is required to break the stick now? With just a flick of your finger the stick breaks. Enzymes, like this press, are used to lower the amount of energy required to initiate a chemical reaction. Regarding enzyme function, consider the following points:

1. Enzymes are not used up or consumed during the chemical reaction. In other words, a single enzyme can serve as a catalyst for multiple reactions.
2. Enzymes are quite specific, so a single enzyme is able to catalyze a reaction between certain reactants (substrates) but not others, which is why we need so many different enzymes. An example that you are familiar with is converting a common disaccharide, sucrose, to two monosaccharides, glucose and fructose. The enzyme that is involved in this reaction would be unable to convert the disaccharide lactose to the monosaccharides glucose and galactose.
3. Enzymes are often named for the substrates on which they act. Thus, the enzymes involved in the reactions above would be sucrase and lactase respectively. Notice that the suffix –ase is added to the name of the substrate.
4. An enzyme's shape governs its function. Each enzyme has an active site where only certain molecules (substrates) can bind. When the substrates bind to the active sites, the enzymes catalyze the chemical reaction, and they are released as a new product.
5. Enzymes are sensitive to changes in temperature and pH. One way to speed up chemical reactions is to turn up the heat, but increasing temperature too much can alter or even destroy cells. Enzymes in the human body function optimally between 35–40° C (95–104° F). They also function best at around a neutral pH level, with a range typically between six and eight. If we change the temperature or pH to values outside the optimum, the enzymes may change shape and lose their function.
6. Enzymes may require "helper" substances to catalyze chemical reactions. These helpers are termed cofactors or coenzymes. Cofactors are inorganic substances such as zinc or iron. Coenzymes are organic molecules like vitamins.
7. Enzymes are not used up or consumed during the chemical reaction. In other words, a single enzyme can serve as a catalyst for multiple reactions. In the example above, the matchmaker could go on to set up other dates between other young men and women.
8. Enzymes are quite specific, so a single enzyme is able to catalyze a reaction between certain reactants (substrates) but not others, which is why we need so many different enzymes. An example that you are familiar with is converting a common disaccharide, sucrose, to two monosaccharides, glucose and fructose. The enzyme that is involved in this reaction would be unable to convert the disaccharide lactose to the monosaccharides glucose and galactose. Our matchmaker above is great at setting up dates between young men and women but is completely useless in helping dogs find their true love!
9. Enzymes are often named for the substrates on which they act. Thus, the enzymes involved in the reactions above would be sucrase and lactase respectively. Notice that the suffix –ase is added to the name of the substrate.
10. An enzyme's shape governs its function. Each enzyme has an active site where only certain molecules (substrates) can bind. When the substrates bind to the active sites, the enzymes catalyze the chemical reaction, and they are released as a new product.
11. Enzymes are sensitive to changes in temperature and pH. One way to speed up chemical reactions is to turn up the heat, but increasing temperature too much can alter or even destroy cells. Enzymes in the human body function optimally between 35–40° C (95–104° F). They also function best at around a neutral pH level, with a range typically between six and eight. If we change the temperature or pH to values outside the optimum, the enzymes may change shape and lose their function.
12. Enzymes may require "helper" substances to catalyze chemical reactions. These helpers are termed cofactors or coenzymes. Cofactors are inorganic substances such as zinc or iron. Coenzymes are organic molecules like vitamins.

4.0

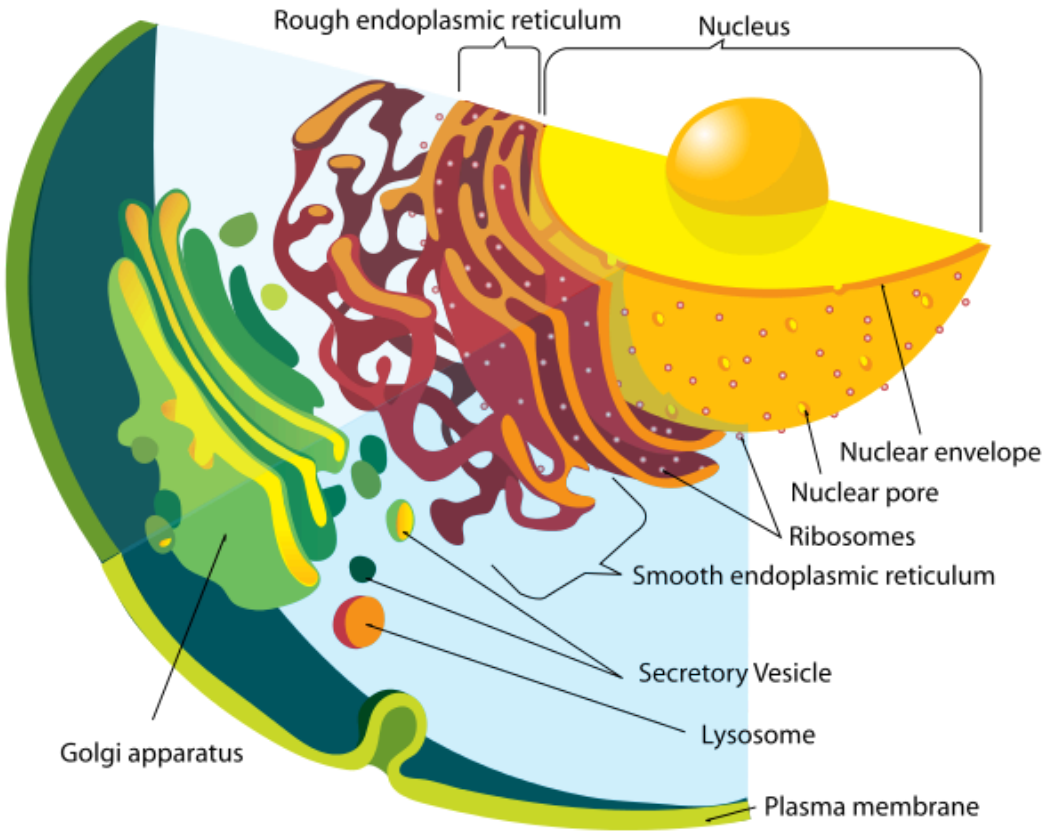
MODULE 4: THE CELL

4.1

CELL STRUCTURES

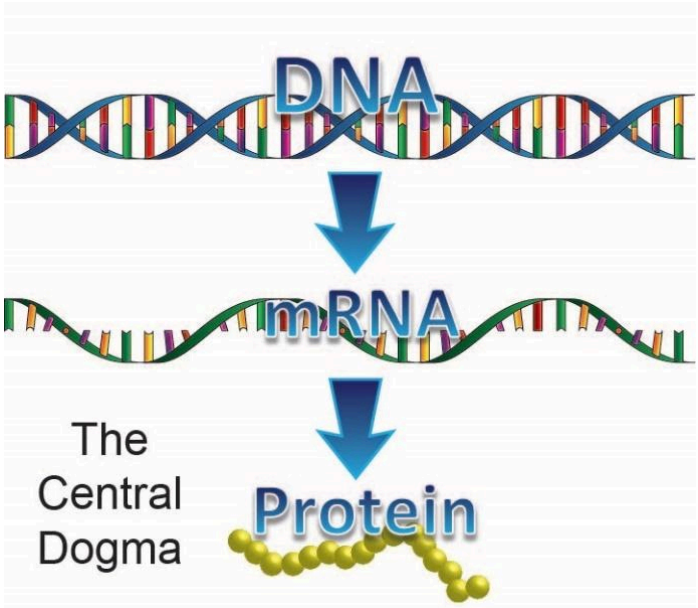
An Interactive Image that allows you to explore the anatomy of a cell has been added to help you. You can see images of many of the bolded terms below. Be prepared to know where these fundamental cell parts are and what they do. As technology progressed to the point of peering deeper and deeper into the world of the cell, it became apparent that the cell cytoplasm, when viewed under a light microscope, was full of even smaller intracellular structures. The structures are collectively called cellular organelles. This introduction will identify these organelles briefly, and subsequent sections will add details.

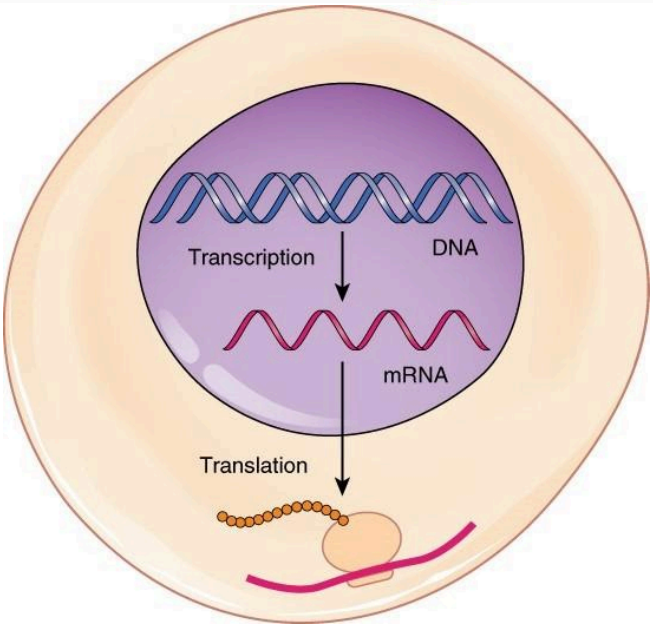
To help illustrate the function of many of these organelles, let us consider the secretion of insulin by beta cells in the pancreas. In order to secrete insulin, the cell must first make it. This process starts in the cell **nucleus**. The nucleus houses the genetic material (DNA) of a human cell and provides a location for **DNA transcription** (the copying of DNA). Importantly, the nucleus is surrounded by two distinct lipid bilayer membranes. The outer membrane belongs to the **endomembrane system** (made up of the nuclear envelope, the endoplasmic reticulum, the golgi apparatus, lysosomes, the plasma membrane, and most vacuoles and vesicles).



Endomembrane System Diagram. Title: File: Endomembrane system diagram en.svg; Mariana Ruiz LadyofHats;Site: http://en.wikipedia.org/wiki/File:Endomembrane_system_diagram_en.svg; License: Public Domain.

The production and secretion of insulin helps illustrate the coordinated efforts of the organelles in this system. Within the nucleus, the insulin gene (located on chromosome 11) is transcribed from DNA to RNA and then further processed into messenger RNA or mRNA. This mRNA is then transported out of the nucleus to **ribosomes** docked to the surface of the **endoplasmic reticulum** (ER). The ER is actually divided into two components: the rough ER and the smooth ER. The rough ER is named "rough" because it is studded with ribosomes, which create a bumpy surface when viewed under an electron microscope. The function of the ribosome is to perform **translation** (the use of mRNA as a template to synthesize protein). The ribosome is specifically suited to interpret the mRNA nucleotide acid code (a series of adenosines, uracils, guanidines, and cytosines—abbreviated A, U, G, and C respectively) and assign the appropriate amino acid in the creation of a polypeptide chain. This is the first step of making a protein. The process of DNA to RNA to protein is called the **central dogma** of biology.





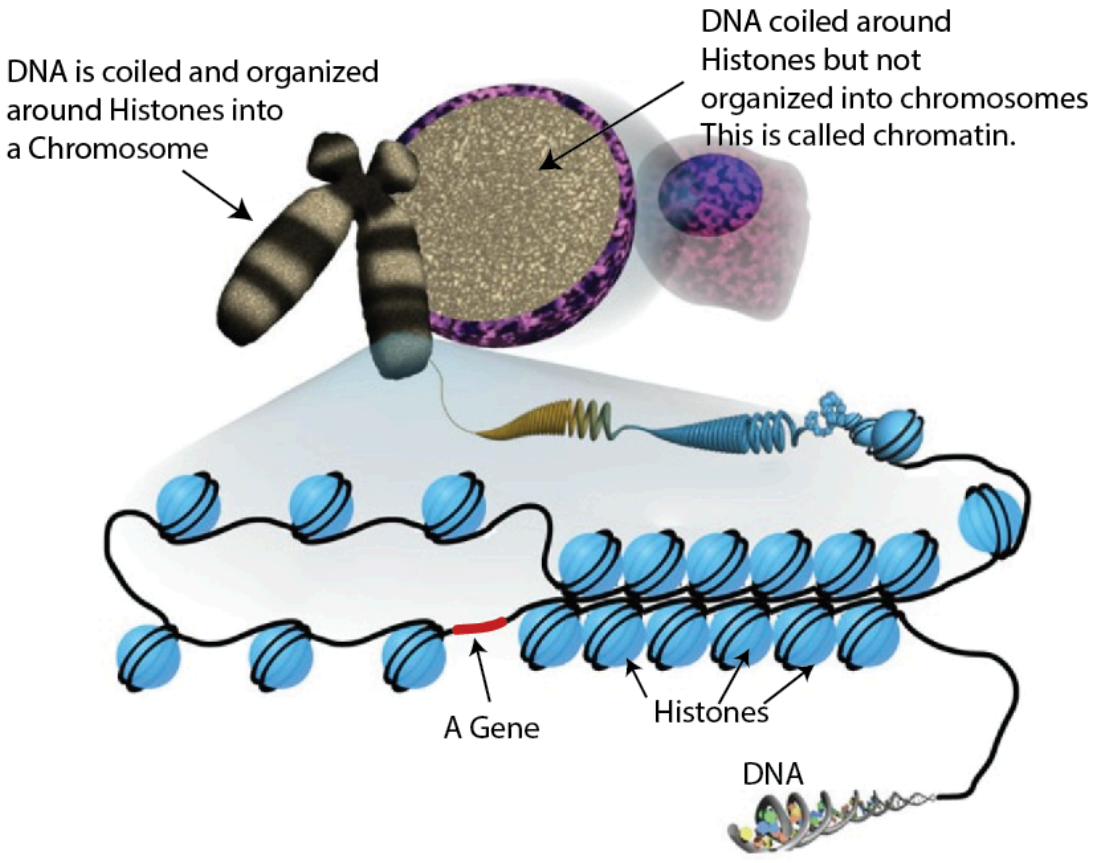
Central Dogma of Biology - DNA Transcription to Translation. File:0328 Transcription-translation Summary.jpg;Author: OpenStax College; Site:http://commons.wikimedia.org/wiki/File:0328_Transcription-translation_Summary.jpg;License:

Within the rough ER, the nascent (immature) insulin protein is folded into primary, secondary, and teriary structures. It is then transported to the **Golgi apparatus**. The Golgi apparatus is the location for processing and sorting (think of a giant UPS mail warehouse). Within the Golgi, the nascent insulin is further processed into mature (functional) insulin and packaged into secretory vesicles. These vesicles (now full of insulin) bud off of the Golgi and are transported, via microtubules, to the **plasma membrane** where they await the proper signal for secretion. Secretion occurs as the vesicle fuses with the **plasma membrane**, expelling its contents into the extracellular space in a process known as exocytosis. Now that you understand the process of how cells can create and secrete insulin let us now examine the roles of each of these organelles in greater detail.

4.1.1

The Cell Nucleus

The nucleus is surrounded by a double membrane bound structure that serves to isolate the nuclear contents from the cellular cytoplasm. This nuclear envelope is dispersed during mitosis and meiosis as the cell prepares to divide. The outer membrane is continuous with the membranes of the rough endoplasmic reticulum. The inner membrane makes the border to isolate the nucleus. The space between the two membranes is continuous with the space (lumen) inside the endoplasmic reticulum, except at various points where the two membranes are connected by specialized structures known as nuclear pores. The nuclear pores serve as transport pathways between the interior of the nucleus and the cytoplasm. The nucleus contains the genetic material (genes) that are organized into long double stranded molecules called DNA. DNA are tightly bound to proteins called histones to form chromatin, which is finally organized into chromosomes.



DNA Structure: Gene, Histones, Chromatin & Chromosomes. Modified image - Title: File: Sha-Boyer-Fig1-CCBy3.0.jpg; Author: unknown; Site: https://books.byui.edu/-FjMn;*License: This work is licensed under the Creative Commons Attribution 3.0License.*

The nucleolus is a region of the nucleus responsible for the synthesis of ribosomes. This region is mad of DNA, RNA, and proteins. Gene messages are copied from the DNA as single strands of RNA, which are further processed into what we call “messenger RNA” (or mRNA) and sent out of the nucleus through the nuclear pores. The mRNA interacts with ribosomes to produce a specific protein. Many genetic mutations result in errors, making the associated proteins non-functional. The nucleus then functions to maintain the integrity of the genes as well as control the turning on or off of the genes. The genes, in turn, regulate the activity of the cells. Thus, the nucleus is the control center of the cell.

4.1.2

The Endoplasmic Reticulum

As mentioned previously, the endoplasmic reticulum has a rough component and a smooth component. The rough endoplasmic reticulum is associated with ribosomes that constantly bind and unbind to the membrane. Ribosomes bind to the endoplasmic reticulum after they interact with an mRNA strand from the nucleus. The ribosomes "read" the mRNA strand and produce the specific protein associated with the code and secrete it into the lumen of the rough endoplasmic reticulum. The newly produced proteins are then folded and prepared for transport to the Golgi complex where they will complete processing prior to being utilized outside of the cell. The smooth endoplasmic reticulum synthesizes lipids, phospholipids, and steroids. In addition, it aids in the breakdown of carbohydrates and steroids. The membrane contains proteins that move Ca^{++} into the structure for storage and thus plays an important role in regulating cellular calcium ion concentrations.

4.1.3

The Golgi Apparatus

The Golgi apparatus is named after the person that discovered it, an Italian physician named Camillo Golgi who discovered the organelle in 1897. Similar to the endoplasmic reticulum, the Golgi apparatus is an organized structure of phospholipid membrane. The compartments of the Golgi body are involved in further processing of proteins that were first made in the rough endoplasmic reticulum. Membrane bound vesicles that arise from the Golgi are distributed to various locations within the cell. The Golgi apparatus is particularly important in the processing of proteins that are destined for secretion outside of the cell. Proteins are sent to the Golgi apparatus from the rough endoplasmic reticulum through transport vesicles that move on the "highway" network of the cell, the cytoskeleton (discussed below). The Golgi apparatus is primarily associated with proteins but also serves in the transport of lipids around the cell and the creation of lysosomes. Perhaps the best analogy for the Golgi apparatus would be that of the post office of the cell.

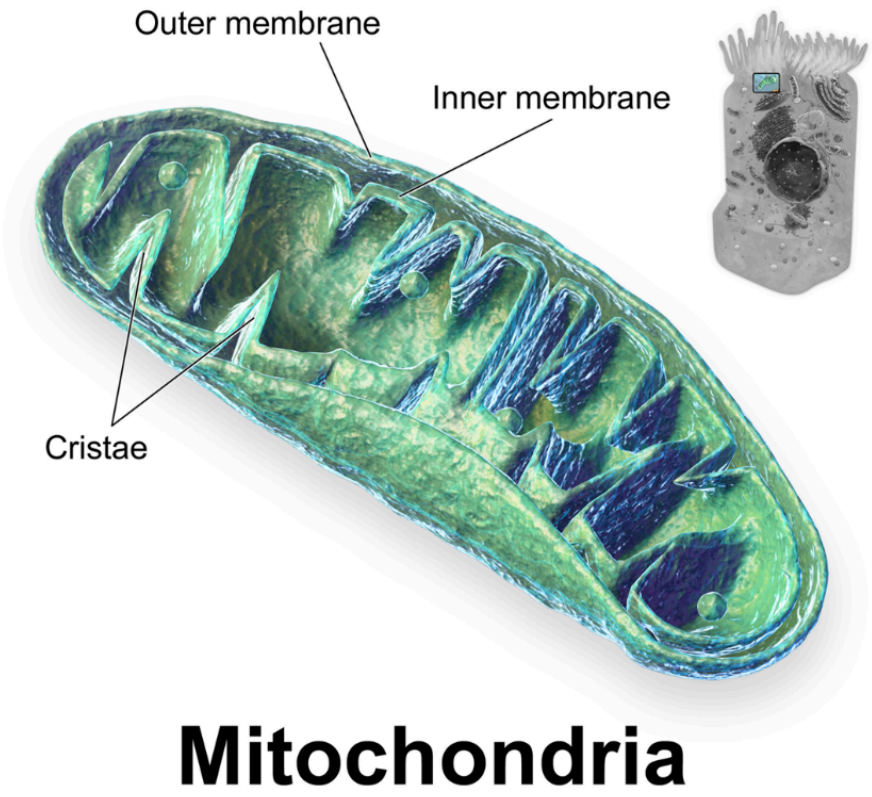
4.1.4

The Mitochondrion

The mitochondrion (or mitochondria in the plural form) is usually described as the “power plant” because it generates the energy (in the form of adenosine triphosphate or ATP) required for normal cellular function. Like the nucleus, mitochondria also have two membranes, which are critical for its function in energy production (further detail will be given later as we study metabolism). Also, like the nucleus, the mitochondrion has its own set of DNA with its own set of genes. Mitochondria are very important in metabolism and they can be stimulated to synthesize proteins from its DNA and even divide and create more mitochondria when metabolic demands increase (this happens with exercise training). The mitochondrion is composed of an outer and an inner membrane (a balloon within a balloon) that gives five distinct structural components.

1. The outer mitochondrial membrane
2. The intramembranous space (space between outer and inner membranes)
3. The inner mitochondrial membrane
4. Cristae (foldings of the inner membrane)
5. The matrix (space of the interior of the mitochondrion)

Each region is associated with a particular function as it relates to mitochondrial activity. The number of mitochondria per cell varies widely with more than 2000 per cell in liver cells and down to zero for red blood cells.



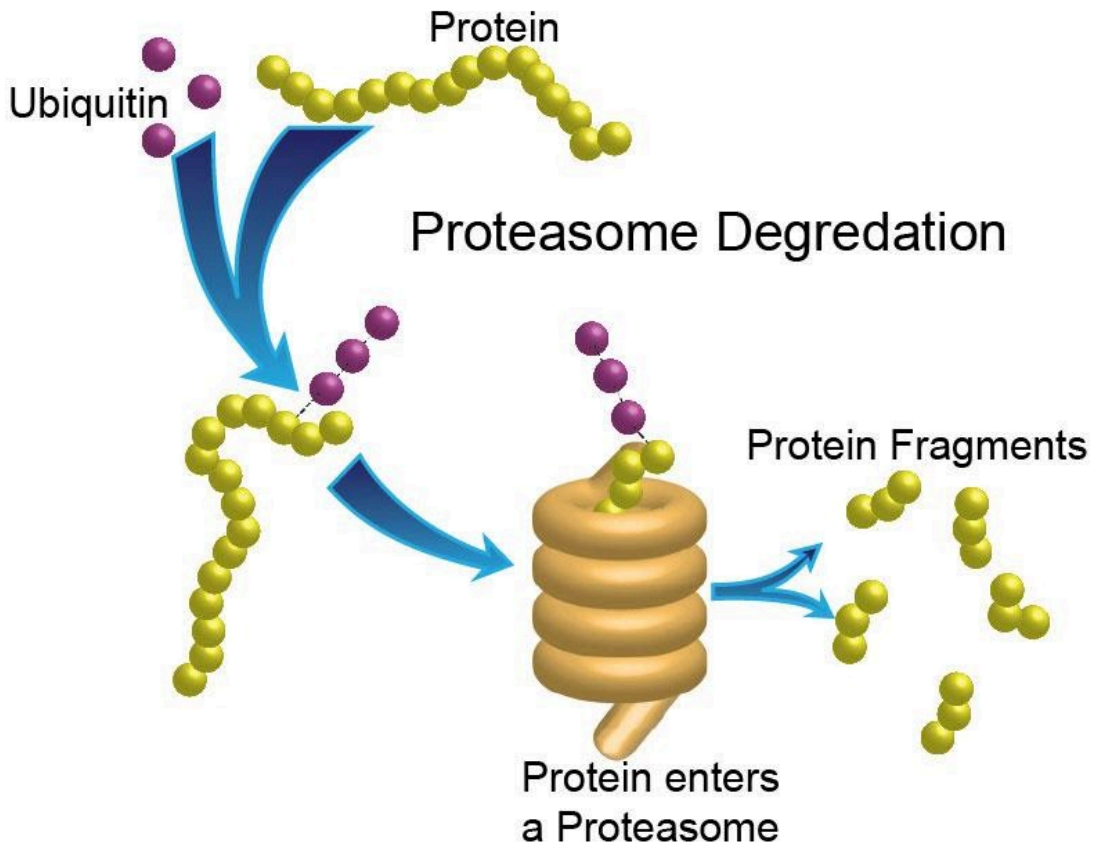
Mitochondria

Mitochondria. File: Blausen 0644 Mitochondria.png; Author: Blausen.com staff. "Blausen gallery 2014". *Wikiversity Journal of Medicine. DOI:10.15347/wjm/2014.010. ISSN 20018762; License: Creative Commons Attribution 3.0 Unported license.*
Site: https://books.byui.edu/-Zaw;

4.1.5

Lysosomes, Proteasomes, and Peroxisomes

As mentioned, **lysosomes** are also part of the endomembrane system. Lysosomes are specialized vesicles that bud off of the Golgi apparatus. A lysosome uses a pump within its membrane to transport high concentrations of H^+ into its lumen, thus lowering the internal pH. The acidic environment of the lysosome allows it to break down macromolecules (such as proteins). Other organelles involved in recycling used or unneeded materials include **proteasomes** and **peroxisomes**. When a cell wants to quickly reduce the amount of a given protein, it can tag that protein with a specific signal (called ubiquitin) that sends that protein to the proteasome for degradation. The peroxisome is responsible for detoxifying harmful substances that may enter the cell by generation of hydrogen peroxide (H_2O_2). Perosisoes are also involved in some metabolic reactions.



Proteasome Degradation. *Original image drawn by BYU-I Biology Department Jan 2015*

4.1.6

The Cytoskeleton

The cytoskeleton, as the name implies, is the structural component of the cell and is composed of a network of proteins that are constantly destroyed, renewed, and newly built. The cytoskeleton functions in maintaining the cell shape, resisting deformation, movement both inside (transport of vesicles within) and migratory movement, cell signaling, endocytosis and exocytosis, and cell division. The cytoskeleton is composed of three major filaments: microfilaments, intermediate filaments, and microtubules filaments. You can explore these components visually at this link:

Cytoskeletal Networks: https://books.byui.edu/-pBri

Video: Cytoskeleton Microtubules: https://youtu.be/5rqbmLiSkpk

Microfilaments are the thinnest of the cellular filaments and are composed of long chains of protein monomers called G-actin. They can generate force by adding monomers that cause the growing strand to push against barriers like the cell membrane. Other proteins, like myosin, can move along the track and pull against it, generating contractile forces in all cells, which is especially important in muscle cells. Intermediate filaments are stronger than micro filaments and thus help maintain the cell shape. The filaments serve as anchors for other organelles; they also serve as cell-to-cell junctions. Intermediate filaments are also used in helping to maintain the shape of the nucleus. Microtubules are the largest of all filaments, with a hollow structure made up of protein monomers called *tubulin*, which wind like a spiraling staircase. Microtubules are closely associated with an organizing center called the **centrosome**. Microtubule networks serve as "highways" for the transport of vesicles and are important for specialized movements like the swirling tail of sperm cells or the flagellum of bacteria. They also play a crucial role during cell division where they function to pull apart and segregate individual chromosomes.

Now that you have learned about the basic structures of each of the cell organelles, you can see how each of these organelles interact with one another by watching this video:

Overview of Cell Structure: https://youtu.be/URUJD5NEXC8

5.0

MODULE 5: CELL MEMBRANES-STRUCTURE AND TRANSPORT

5.1

STRUCTURE OF THE CELL MEMBRANE

One of the challenges faced by all living things, be they amoebae or humans, is to separate their internal environment from the external environment. Critical nutrients must get into the cells, and waste must get out. To make matters more complex, cells need to be able to regulate that movement, letting the materials cross sometimes and preventing them from crossing at others. Another challenge is finding a way for cells to communicate with each other. Cells in the brain, for example, need to be able to tell cells in the heart to beat faster. The solution to these challenges lies in the properties of the cell membrane (also called the **plasma membrane**). This delicate structure is essential to the life of cells. When the membrane loses its ability to carry out these processes, the cell dies.

In this lesson, we will study this amazing structure. We will learn how it allows some things to readily cross and prevents others. Hopefully, you will gain an appreciation of its complexity and come to realize how important it is to cellular function.

5.1.1

Fluid Mosaic Model of the Membrane

The plasma membrane is more than just a sack to hold the contents of the cell. It plays an important role in cellular function and the maintenance of homeostasis. One obvious function is to regulate what enters and leaves the cell. This process is highly coordinated and very specific. In addition, the cell membrane responds to countless chemical messengers in ways that alter the activity of the cell. As we discuss the structure of the plasma membrane, keep in mind that this description also applies to other membranes that are components of intracellular organelles.

Our modern model of the cell membrane is called the **Fluid Mosaic Model** *of the Cell Membrane*. The word *fluid* implies that the membrane is constantly changing and moving. Indeed, it is not a static structure but one that changes depending on cellular need and environment. A good example of this fluidity can be seen with the uptake of glucose into muscle cells. The plasma membrane of muscle cells is normally impermeable to glucose, preventing it from entering the cell. Only when a signal molecule (insulin) is present can glucose enter. The presence of this signal results in the insertion of special glucose transporters into the membrane, allowing glucose to enter the cell. When insulin is no longer there, the carriers are removed, demonstrating the ability of the membrane to change depending on stimulus of the cell. Additionally, components of the membrane are not rigidly fixed in one area but often have the freedom to move laterally within the membrane.

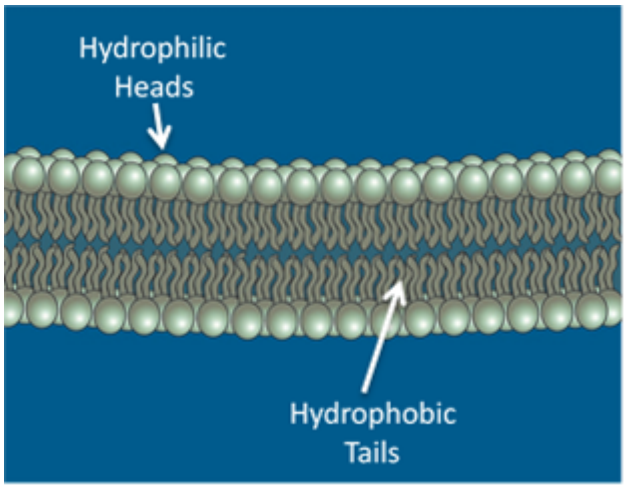
The term *mosaic* conjures up an image of numerous small and different pieces. Indeed, the membrane contains many different components including lipids, proteins, and carbohyrdrates. The following link shows the structure and function of the membrane:

https://books.byui.edu/-tRDY (Transcription Available).

5.1.2

Membrane Phospholipids

A key component of the membrane is a double layer of phospholipids, the **phospholipid bilayer**. This bilayer forms the scaffolding into which the other components of the membrane are housed. Recall phospholipids are composed of a hydrophilic head containing a phosphate group and two hydrophobic tails composed of long chain fatty acids. This bilayer has a central hydrophobic region and two outer hydrophilic sections, one facing the aqueous interior of the cell and one facing the aqueous extracellular space (see figure below).



Phospholipid Bilayer. *Image created by BYU-IU student, Hannah Crowder 2013*
In water, phospholipids can form a bilayer. The hydrophobic fatty acid tails turn away from the water, and the hydrophilic phosphate heads turn towards the water.

The hydrophobic core of the membrane creates a barrier, preventing hydrophilic substances, such as ions and large polar molecules, from moving across the membrane. Hydrophobic (lipid soluble or lipophilic) materials, on the other hand, typically move readily across the membrane. Because some things easily pass through the membrane and others do not, we describe the membrane as being **selectively permeable**. The following link may help you better understand the concept of selective permeability:

https://books.byui.edu/-Waot (Transcription Available)

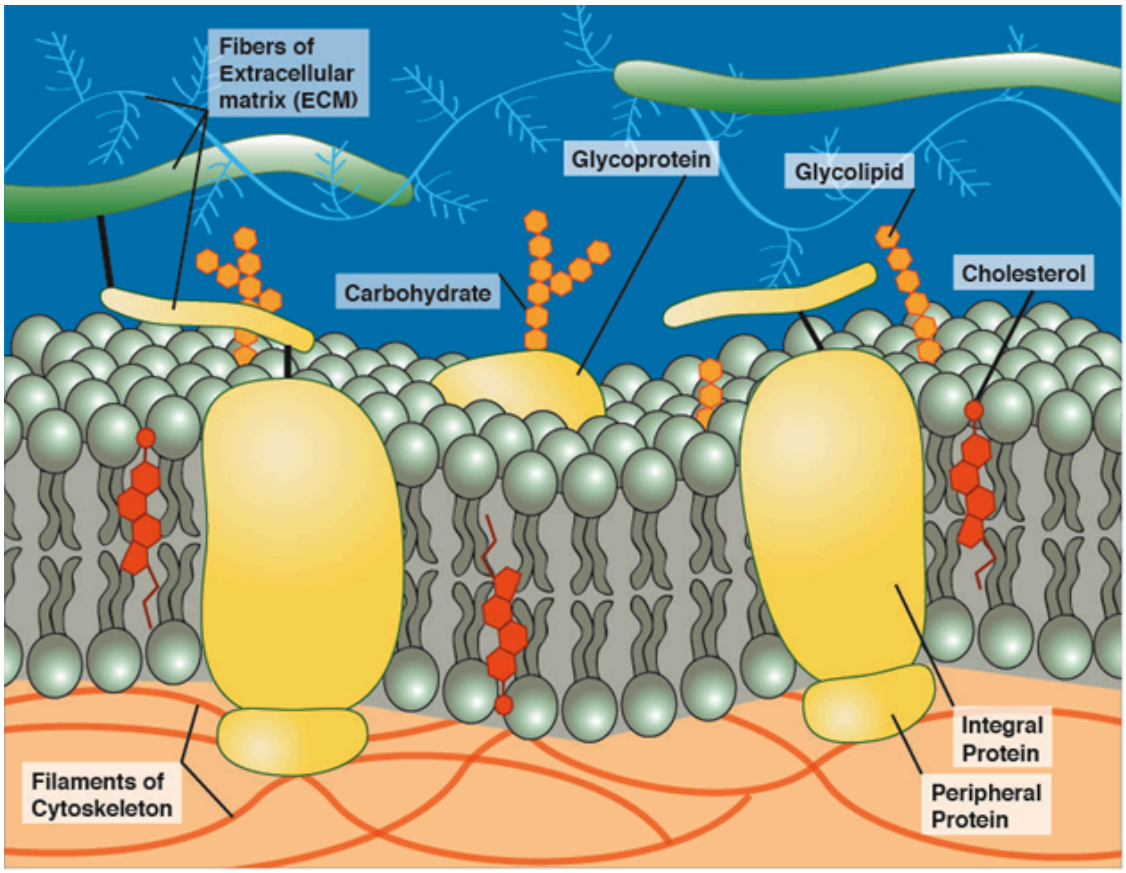
In addition to the phospholipids, another important lipid found in membranes is cholesterol. Cholesterol is a hydrophobic molecule and resides among the fatty acid tails of the phospholipid bilayer. As mentioned above, the membrane exhibits fluidity, allowing movement of components within the membrane. Cholesterol plays an important role in regulating the fluidity of the membrane across a range of temperatures the body is exposed to. While it is true that our core body temperature remains fairly constant, temperatures in our extremities may vary considerably. Think of the range of temperatures the cells in your hands are exposed to. At high temperatures, cholesterol enhances the interactions between phospholipid fatty acids and prevents destabilization and melting of the membrane. At low

temperatures, cholesterol prevents phospholipid tail groups from interacting too strongly with each other, a condition which would stiffening the membrane and decrease fluidity. Thus, without cholesterol the membrane might be compromised leading to impaired cellular function. Together, phospholipids and cholesterol comprise nearly 50% of the membrane.

5.1.3

Membrane Proteins

Making up about another 50% of the membrane are the membrane proteins. The figure below demonstrates the relationship of the membrane proteins with the phospholipid bilayer. Note that some of the proteins are found only on the inner or outer surface of the membrane. These are called **peripheral or extrinsic proteins** because they do not extend through the membrane. One function of the peripheral proteins is to attach the membrane to the cytoskeletal proteins inside the cell or to proteins of the extracellular matrix. For example, the cells lining the blood vessels utilize peripheral proteins to attach to the tissues outside the vessel, thus holding the cells in place.

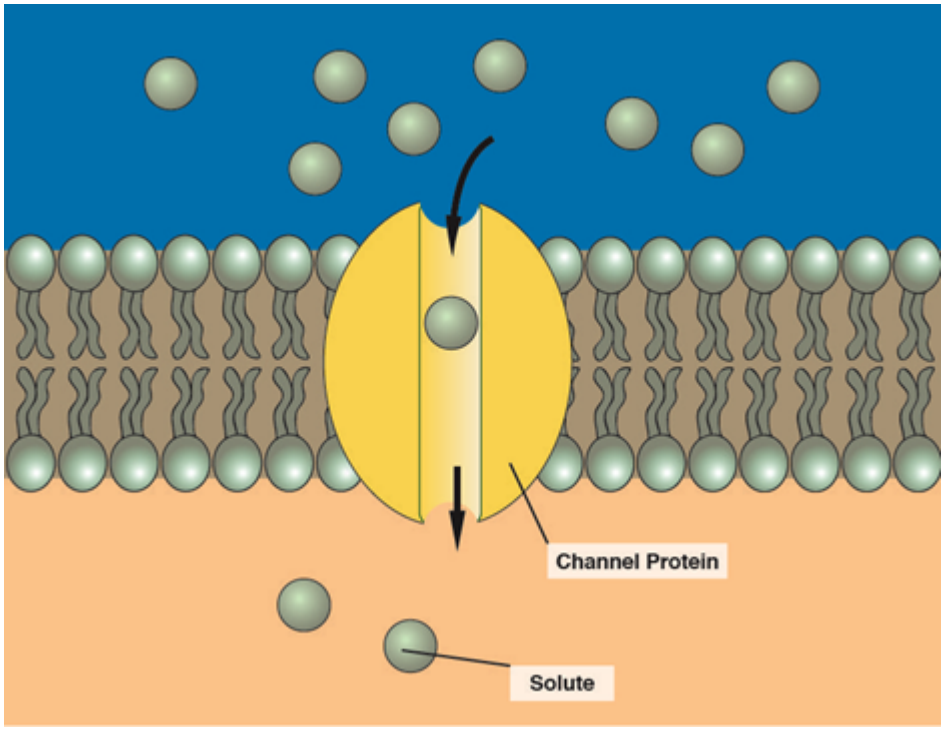


Cell Membrane Model: Relationship of Lipids, Proteins & Carbohydrates. *Image created by BYU student, Hannah Crowder 2013*

Other proteins pass all the way through the membrane. These proteins are called **integral or intrinsic proteins** and have segments that associate with the hydrophobic region of the membrane. These integral proteins perform a number of important functions in the cell. Based on their functions, these integral proteins can be grouped into the following categories:

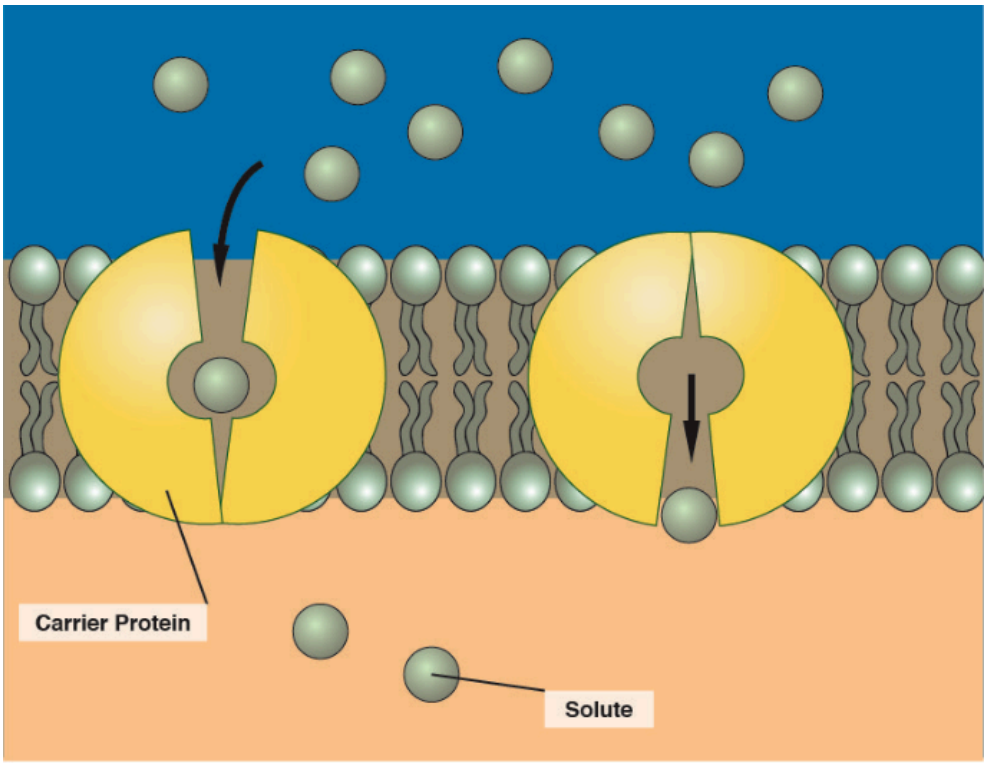
Transport Proteins

Integral proteins can act as transporters that facilitate the movement of compounds across the membrane. One type of transport protein, called **channels**, form a 'tunnel' for hydrophilic materials, such as ions and even water to cross the membrane. These channel proteins are usually gated; like a door, they allow substances to cross only when they are open. We will have more to say about channel gating later.



Channel proteins allow solutes, such as ions, to move across the membrane. *Image created by BYU student, Hannah Crowder 2013*

Carrier proteins are another type of transport protein. Carriers have sites that bind to specific solutes. For example, one type of carrier binds with glucose, while another carrier binds to urea. Once the solute binds, the carrier protein changes shape, allowing the solute to move across the membrane. Imagine a revolving door. As these doors turn (change shape), they are open to either the inside of the building or to the outside but never to both at the same time. You can enter a revolving door from the outside of a room and move the door until it is open to the inside of the room. At no time in this process was the door open to both sides at the same time. This is how carrier proteins work. Carrier proteins bind to solutes and then move them across the membrane by changing shape.



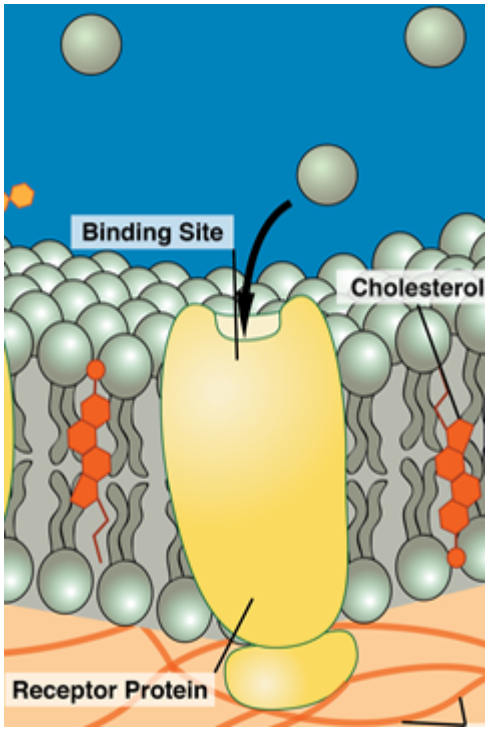
Carrier Proteins. *Image created by BYU student, Hannah Crowder 2013*

Enzymes

Integral membrane proteins can function as enzymes, catalyzing important chemical reactions. The enzyme, lactase, which digests the disaccharide lactose in the small intestine is an integral membrane protein in the cells that line the lumen of the duodenum. The discomfort associated with lactose intolerance is caused by having insufficient amounts of this enzyme in the body.

Receptor Proteins

Integral proteins may act as receptor proteins and allow the cell to respond to extracellular chemical messengers which regulate the activity of the cell. When a chemical signal (also known as a ligand) binds to its specific receptor protein, it transmits a signal to the inside of the cell through a shape change in its transmembrane protein structure. This shape change will then activate or inhibit intracellular events that result in altered cell function. For example, epinephrine (adrenaline) is a ligand that binds to receptors on specialized cardiac cells causing intracellular changes that make your heart beat faster when you are frightened or experiencing an 'adrenaline rush.' There are many types of receptor proteins expressed in our bodies, but we will look at one of the most abundant and well-studied: G-protein coupled receptors.

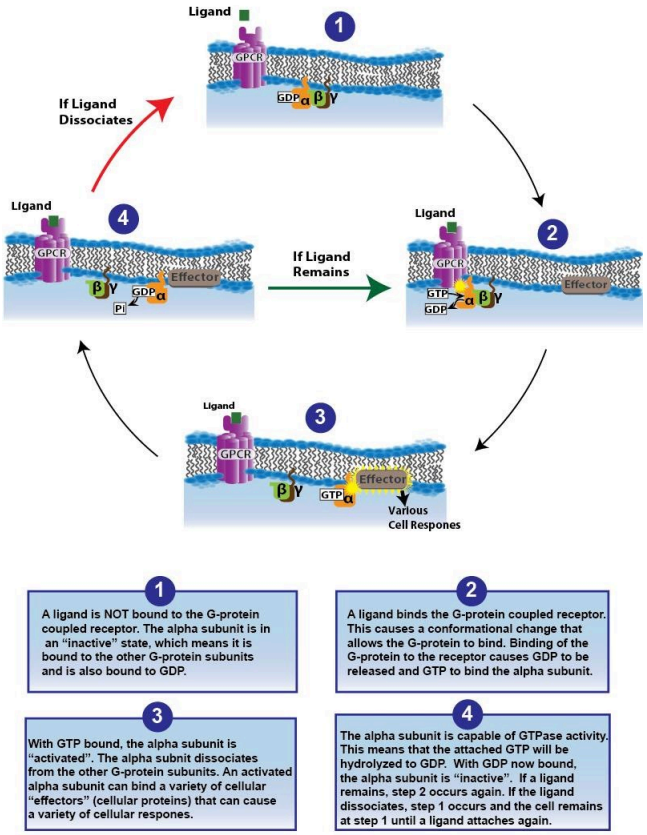


Receptor Protein. *Image created by BYU student, Hannah Crowder 2013*

G protein-Coupled Receptor (GPCR)

The GPCR complex is composed of two units: a receptor protein that binds to the chemical signal (the ligand) and the G protein complex associated with the inner side of the membrane (i.e. a peripheral protein complex). The GPCR has a ligand binding site on the external surface and a G protein binding site on the internal surface. The G protein complex is composed of three subunits: the alpha, beta, and gamma subunits. The alpha subunit has a site that can bind **Guanosine Triphosphate (GTP)** or **Guanosine Diphosphate (GDP)**, hence the name ***G* protein**. In its inactive form, the G-alpha subunit is bound to GDP, and the three subunits (alpha, beta, and gamma) are bound together. When a ligand binds to the receptor on the surface of the cell, the G protein binding site changes shape, allowing the G protein to bind to the intracellular region of the receptor. This binding causes the G protein to then change shape, and the GDP exits the binding site on the alpha subunit and is replaced by a GTP from the cytoplasm. The binding of GTP causes the alpha subunit to separate from the other two subunits (beta/gamma dimer). Once separated, the alpha subunit (and sometimes the beta/gamma dimer) can then bind to and activate other proteins inside the cell. The mechanism of action is typically mediated by one of two enzymes: adenylate cyclase or phospholipase C. These effector proteins will be discussed later in the semester, but for now remember that they create intracellular signaling molecules called 'second messengers' that result in changes in cell function. Cellular responses include activation of metabolic enzymes, opening or closing ion channels, turning on transporters, initiating gene transcription, regulating motility, regulating contractility, stimulating secretion, and even controlling memory. After a short period of time, the G-alpha subunit hydrolyzes the GTP into a GDP and phosphate, allowing it to reunite with the beta/gamma dimer, turning off the signal.

To date, approximately 800 genes for G protein-coupled receptors have been identified. G-proteins are very common in physiology, and it is important to study the details related to these receptors. It would be a good idea to learn and be able to explain the following figure.



Ligand Activation and G-Protein Effect. *Image drawn by JS Fall 2014*

Attachment Proteins

Integral proteins are involved in attaching cells to each other, as well as to the extracellular matrix and to intracellular structural proteins. Often, a peripheral protein functions as a link between the integral proteins and the structural proteins or the extracellular matrix. These attachments can confer tissue strength and shape. The inability to form these connections can result in several pathological conditions, including muscular dystrophy.

Marker Proteins

These proteins allow cells to identify one another. Functions of these marker proteins include the ability of sperm cells to recognize the oocyte during fertilization, as well as the ability of our immune cells to distinguish between our own cells and a foreign cell, such as a bacterial cell, that might be trying to invade our bodies.

5.1.4

Carbohydrates

In addition to lipids and proteins, the membranes also contain carbohydrates. These are short-chained polysaccharides (oligosaccharides) that attach to the proteins and lipids on the extracellular layer of the membrane. If attached to a protein, they are called **glycoproteins**, and if attached to a lipid, they are called **glycolipids**. One function of these oligosaccharides bound to membrane proteins or lipids is to form additional cell markers. Your blood type (A, B, AB, or O), for example, is determined by glycoproteins expressed on your red blood cells. Additionally, some cells, such as the apical surface of epithelial cells, have a dense layer of glycoproteins referred to as the **glycocalyx**. The glycocalyx has been implicated in cell recognition during development, adherence of cells to each other, and playing a role in the permeability of the membranes.

5.2

MEMBRANE TRANSPORT

One of the primary functions of the membrane is to separate the intracellular environment from its extracellular environment. This separation is crucial for the maintenance of the proper conditions for cell function. To perform this important function, the membrane must regulate what enters and leaves the cell. For example, the proper nutrients must be allowed to enter, and wastes must be allowed to leave the cell. Additionally, some things must not be permitted entrance to or exit from the cell. In this section, we will discuss how various substances are moved across the plasma membrane.

Passive vs Active Processes

Watch this Video on Passive Diffusion

Watch this Video on Active Diffusion

Processes that move substances across membranes can be grouped into two general categories based on whether the process requires energy or not. If no energy input is required for the transport, we say particles move via **passive transport**. On the other hand, if the process requires cellular energy, then it is an **active transport** process.

Simple Diffusion
Facilitated Diffusion
Active Transport
Osmosis

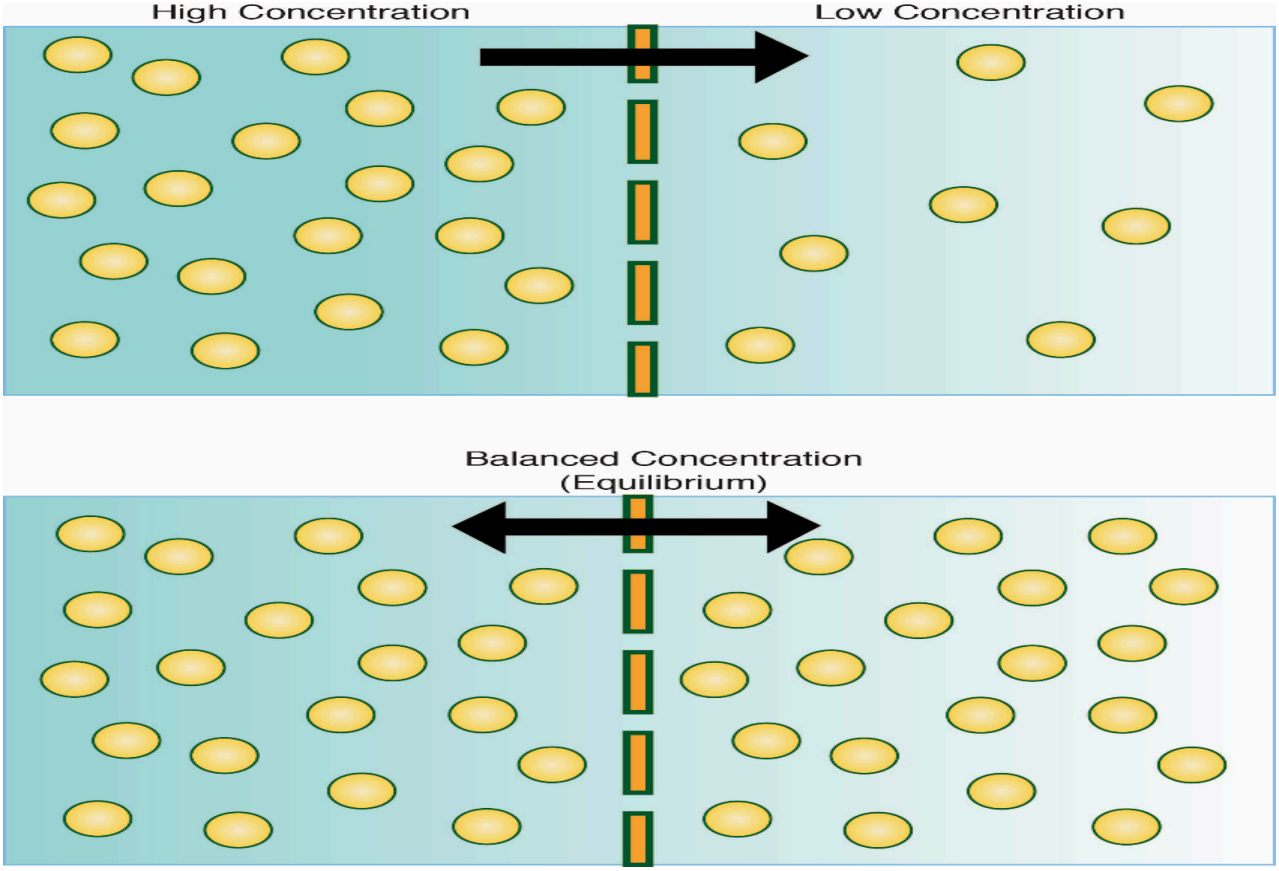
5.2.1

Simple Diffusion

Diffusion is a process that results from the fact that molecules are constantly in a state of random movement. They move in a straight line until they collide with another molecule, causing them to bounce off in a different direction. If there is an initial, unequal distribution of the molecules (i.e. more concentrated in one area than another), the constant random movement and collisions cause them to eventually become equally distributed. This process of gradual movement from where they are more concentrated to where they are less concentrated is called *diffusion*. We refer to the concentration difference as the **concentration gradient**.

Therefore, substances diffuse down their concentration gradients (from high to low concentration). Once the molecules are evenly distributed, we say that we have reached a state of **diffusion equilibrium**, and even though the molecules are still moving, there is no longer any net change in concentration. You can observe this phenomenon by carefully placing a drop of food coloring into a glass of water. The dye gradually moves through the liquid until it is evenly dispersed in the water. In the body, if the material in question can pass through the cell membrane without the aid of a membrane protein, we refer to the process as **simple diffusion**. Solutes that cross the membrane by simple diffusion tend to be hydrophobic. Examples of substances that cross the membrane by simple diffusion are the gasses CO2 and O2. The following video demonstrates the process of diffusion and explores some of the factors that influence the rate of diffusion. See if you can answer the questions posed in the video before reading the next section.

https://books.byui.edu/-bNsc (Transcription Available)



Simple Diffusion: Process of Moving from High to Low Concentration to Reach Equilibrium.
Image created by BYU-Idaho student, Hannah Crowder 2013.
The top panel shows the diffusion of solute from left (high concentration) to the right (low concentration) until an equilibrium is established. Once a diffusion equilibrium exists, there will no longer be any net movement of solute (lower panel).

Factors That Affect the Rate of Diffusion

The rate at which the solute diffuses is affected by several factors.

Concentration gradient: The greater the difference between the concentrations on the two sides of the membrane, the faster the rate of diffusion.

Temperature: The higher the temperature, the faster molecules move. Therefore, as temperature increases, the rate of diffusion increases.

Size of molecule: Smaller molecules tend to travel further before colliding with other molecules, so diffusion rates are faster for smaller molecules.

Viscosity of the medium: The viscosity is a measure of the "thickness" of the solvent. An increase in viscosity decreases the rate of diffusion.

Membrane permeability: Since we are talking about the movement of solutes into and out of the cell, the permeability of the membrane to the solute will affect how fast solutes diffuse across the cell membrane. For example, ions and other charged molecules that are hydrophilic do not readily cross the membrane due to the hydrophobic core of the bilayer. Conversely, oxygen and carbon dioxide, both nonpolar molecules, can readily diffuse through the membrane.

Surface area: The greater the surface area of the membrane, the faster the rate of diffusion is across the membrane. Areas in our bodies, where diffusion is especially important, typically employ structural modifications that increase the available surface area. For example, in the lungs, the hundreds of millions of small alveoli have a total surface area of nearly 70 square meters for gas exchange! This is approximately the same size as a typical two-bedroom apartment in Rexburg.

Distance: Diffusion is quite rapid over short distances but gets slower the further it goes. The time it takes for something to diffuse is proportional to the square of the distance. Therefore, if it takes one second to diffuse one centimeter, it would take 100 seconds to diffuse 10 cm and 10,000 seconds to diffuse 100 cm. So, to go 100 times further takes 10,000 times longer. In the body, diffusion is quite sufficient to cross the thin cell membrane, but to travel long distances by diffusion would be very slow. This is why we have other mechanisms, like the blood circulation and motor proteins along microtubule networks for moving substances long distances.

5.2.2

Facilitated Diffusion

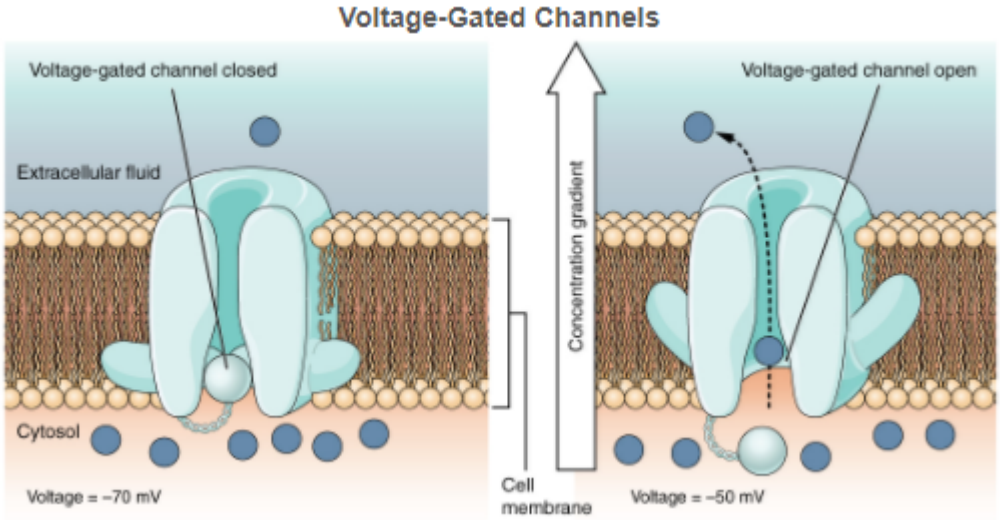
Not all solutes can pass directly through cell membranes. Molecules that are large, or that have an electrical charge, generally are prevented from moving through the membrane. However, many of these solutes need to be able to enter or leave the cell. So, how does the cell solve this dilemma? Recall that embedded in the cell membrane are several types of proteins that pass completely through the membrane (the integral membrane proteins). There are several specialized integral proteins that assist in the diffusion of solutes across the membrane. This type of diffusion is referred to as **facilitated diffusion.** Facilitated diffusion can occur in two different ways, through channel proteins and carrier proteins. Below are two links to videos that demonstrate the properties of these two passive processes that will discuss in more detail below.

https://books.byui.edu/-YRw (Transcription Available)

https://books.byui.edu/-HXs (Transcription Available)

Channel Proteins

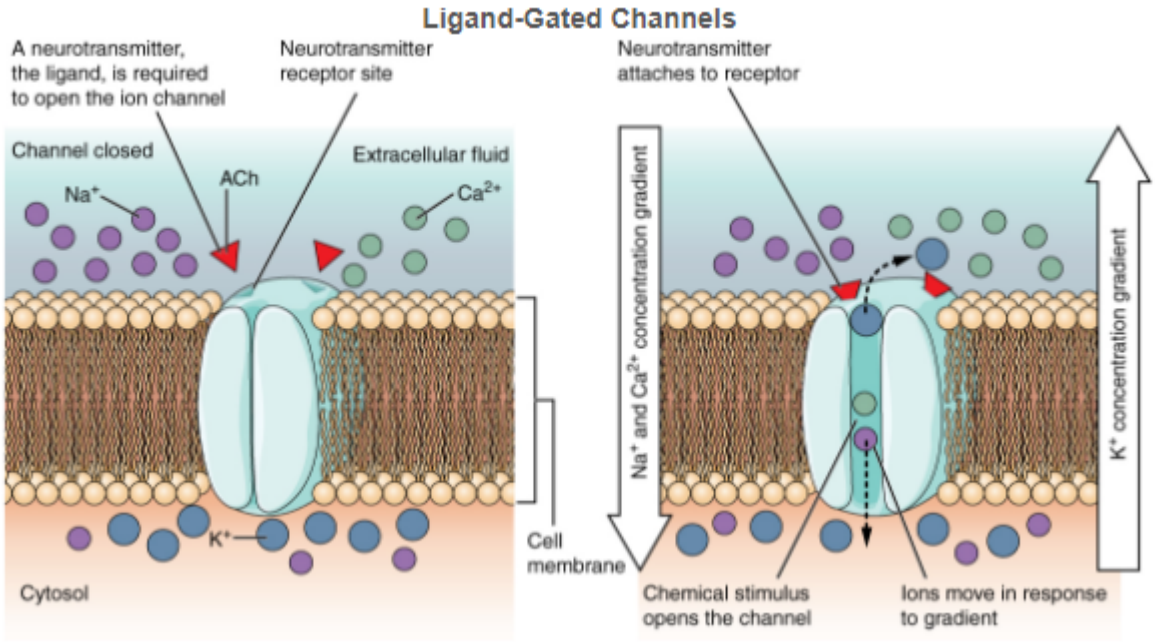
The first is via **channel proteins**. These channel proteins resemble fluid filled tubes through which the solutes can move down their concentration gradients across the membrane. These channels are often responsible for helping ions, such as Na^+, K^+, Ca^{2+}, and Cl^-, cross the membranes. Even though they are open tubes, they often only allow very specific ions to pass through them. For instance, a K^+ channel may allow K^+ to pass through but not Na^+ or Cl^-. Also, as we shall learn later, the regulation of the movement of the various ions across the membranes is crucial for many important cellular functions. These channels, therefore, are often gated (they have doors or gates that can be opened or closed). Depending on the channel, these gates may respond to voltage differences across the membrane (**voltage-gated channels**), specific signal molecules (**ligand-gated channels**), or even stretching or compressing of the membrane (**mechanically-gated channels**).



Voltage Gated Channel. *Author: OpenStax College; Site: https://books.byui.edu/-yThl License: Licensed under a Creative Commons Attribution 4.0 License*

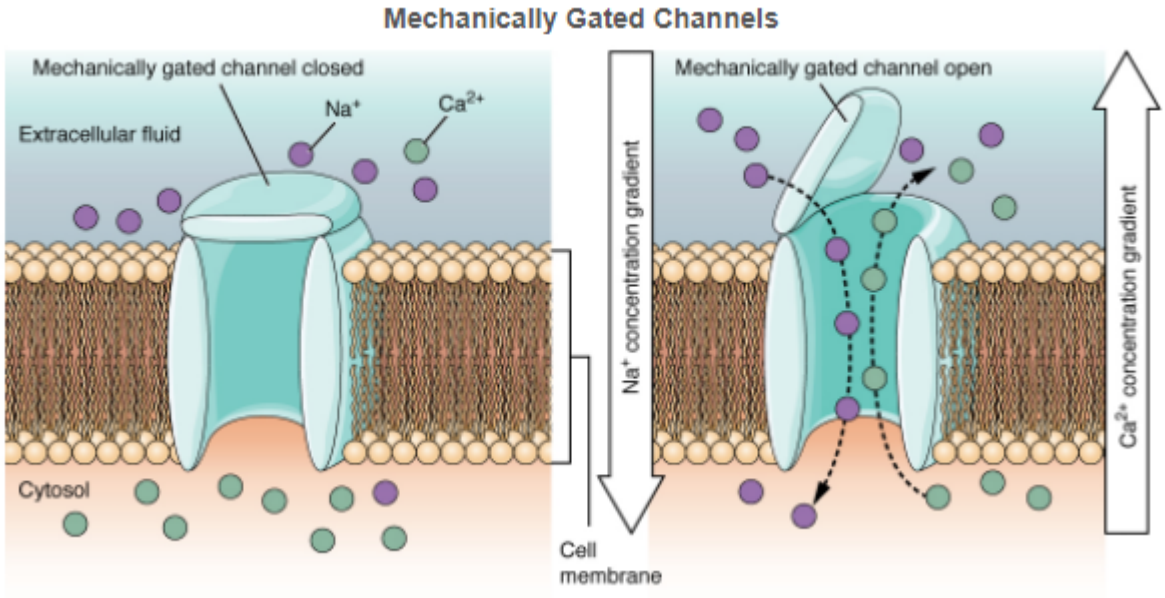
Voltage-gated channels (shown above) open when membrane voltage changes. We will discuss electrophysiology later on in this chapter but understand now that the concentration of ions (charged particles) in the intracellular fluid relative to the extracellular fluid creates a voltage (difference in charge). Amino acids comprising the channel protein are sensitive to changes in voltage which can cause the channel to open for a specific ion.

In ligand-gated channels the pore opens when the ligand binds to a specific location on the extracellular surface of the channel protein. Acetylcholine is the ligand shown in the example below.



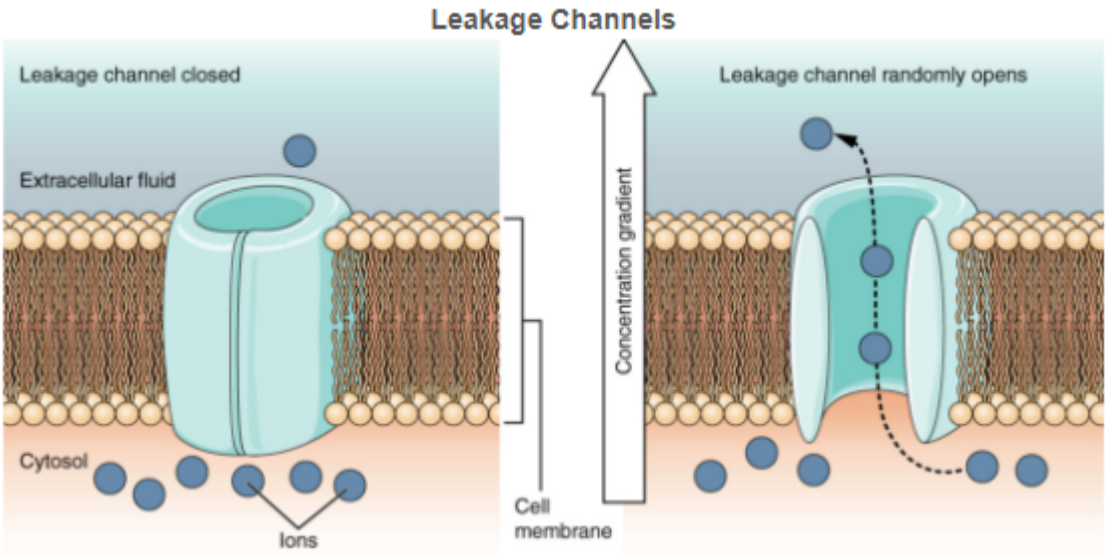
Ligand-Gated Channels. Author: OpenStax College; Site: https://books.byui.edu/-dfk License: Licensed under a Creative Commons Attribution 4.0 License

When a mechanical change happens such as pressure (e.g. touch), or a change in temperature, mechanically-gated channels open.



Mechanical-Gated Channels. Author: OpenStax College; Site: https://books.byui.edu/-ULoF License: Licensed under a Creative Commons Attribution 4.0 License

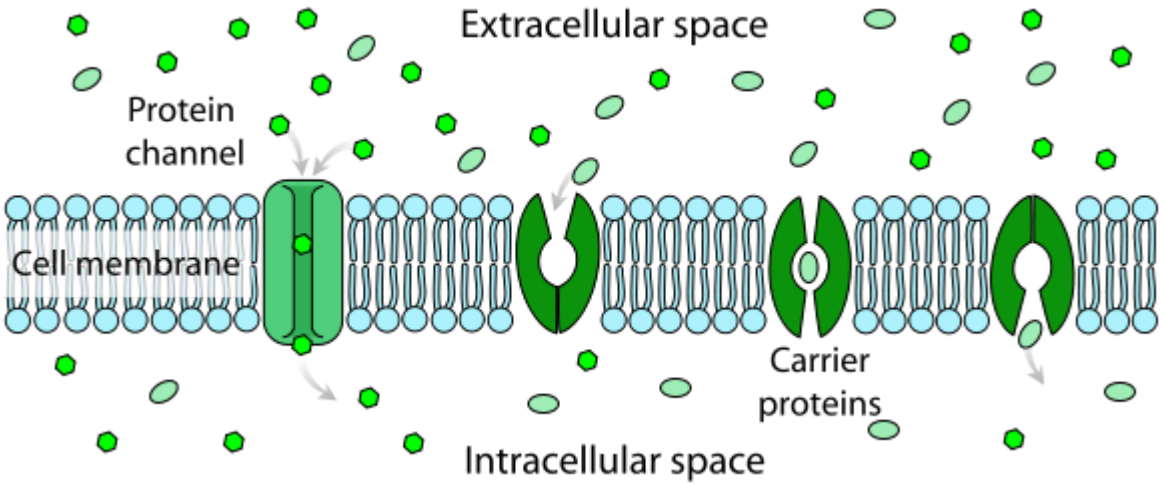
Some channels are not gated and are always open. These channels are called passive or leak channels. Perhaps the most well-known leak channel is the K^+ leak channel which contributes significantly to the resting membrane voltage.



Leak Channel. Author: OpenStax College; Site: https://books.byui.edu/-Mpz License: Licensed under a Creative Commons Attribution 4.0 License

Carrier Proteins

The second type of facilitated diffusion utilizes carrier proteins in the membrane and is known as carrier-mediated transport. Unlike the channel proteins, carriers bind to a specific solute on one side of the membrane which causes the carrier to change shape, allowing solute access to the other side of the membrane (think of a revolving door).

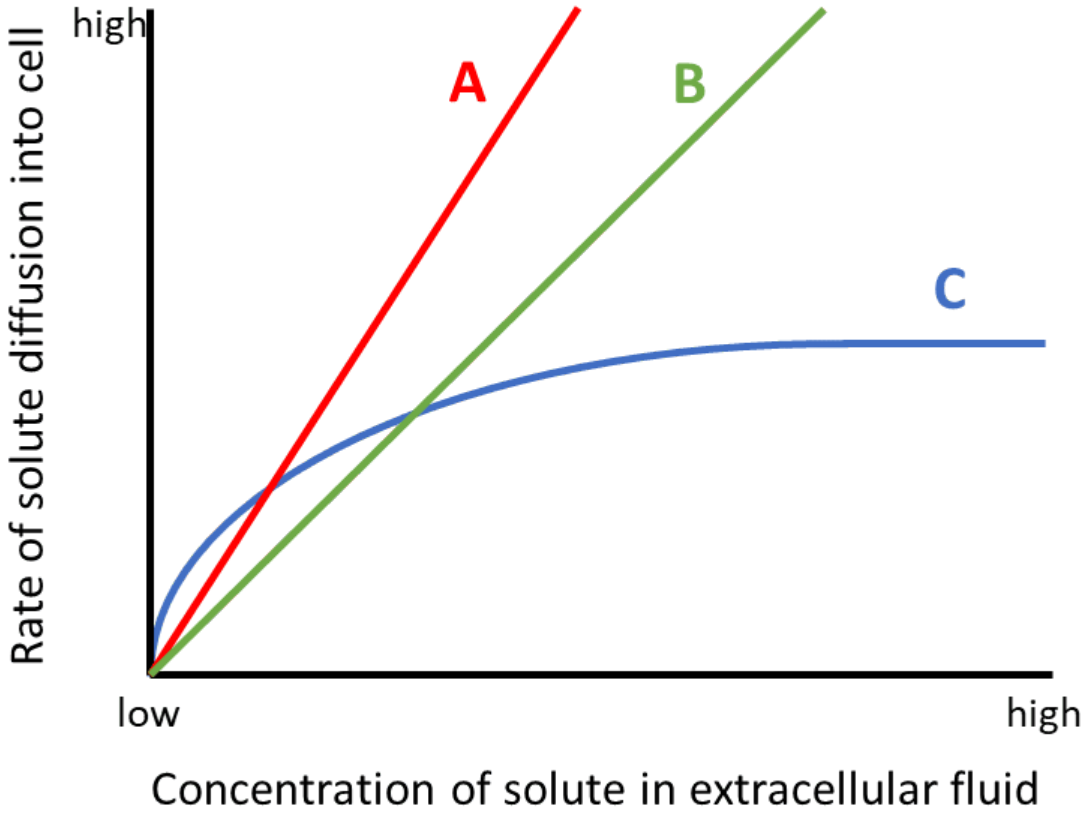


Carrier Proteins. By LadyofHats Mariana Ruiz Villarreal [Public domain], via Wikimedia Commons

Like the channel proteins, these carriers can be very specific for the solute they transport since the solute must bind to a receptor site within the carrier protein before it changes shape. Another interesting characteristic of these carriers is that they have a maximum rate of transport and can thus become saturated if the solute concentration is high enough.

Just as a revolving door can only allow a so many people to enter a building at one time, a carrier protein can only transport a specific amount of solute into the cell in a given time. Here is an example:

An important family of carrier proteins transport glucose across cell membranes. They are called glucose transporters (or GLUTs). There are 12 GLUTs that have been identified. Their distribution and specificity vary. For example, **GLUT2** is found in the kidney, liver, and pancreatic islets while **GLUT4** is found in skeletal muscle and fat tissue. GLUT2 and other glucose carriers expressed in the kidney enable the reabsorption of glucose that gets filtered out of the blood by the kidney. In the case of uncontrolled diabetes where blood glucose levels are super high, the kidney GLUT2 proteins become saturated with glucose and are unable to fully reabsorb the glucose back into the blood. As a result, glucose is lost in the urine, a condition known as glucosuria. Observe the graph below. Read the x and y axis labels. Can you determine which lines represent channel-mediated, carrier-mediated, and simple diffusion?



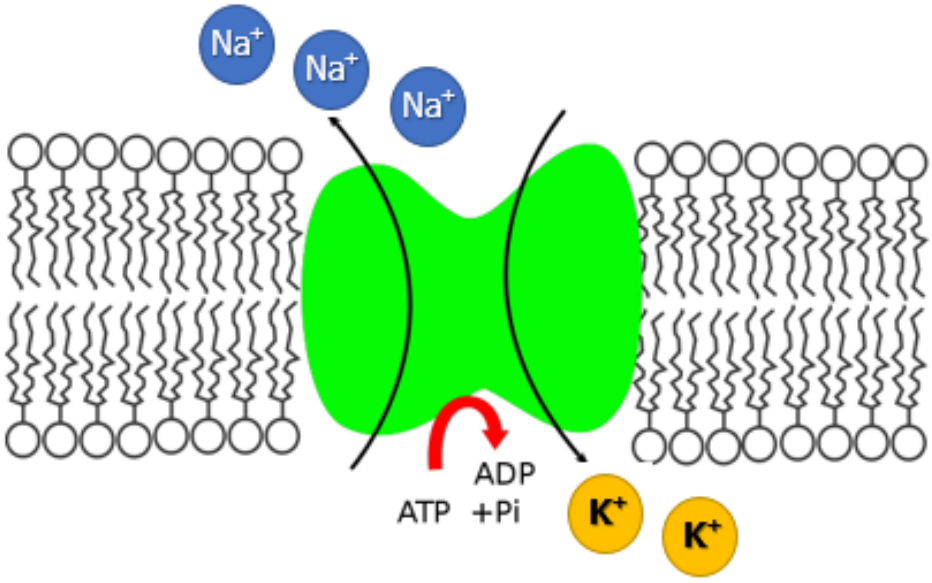
5.2.3

Active Transport

To this point, the transport processes we have discussed have all been passive processes in which the solute movement has been down a concentration gradient with no input of energy required. However, there are times when it is important for the cell to be able to move solutes against their concentration gradient (i.e. moving a solute across the membrane where it is higher in concentration). Just like moving water from the spillway of a dam back to the reservoir, these processes require an energy source and are called **active transport** processes.

Primary Active Transport

Primary active transport requires a carrier protein that is much like the proteins involved in carrier-mediated diffusion mentioned above. However in this case, the carrier has an ATP binding site, which upon hydrolysis into ADP and inorganic phosphate (Pi) provides the energy to move solute against its concentration gradient. These transport systems can move one or multiple ions across the membrane. One of the most important primary active transport proteins is the **Na^+, K^+-ATPase.** This protein moves three sodium ions out of the cell and two potassium ions into the cell for each ATP hydrolyzed. Potassium is the primary intracellular cation in the body while sodium is the primary extracellular cation, and the Na^+, K^+-ATPase is responsible for maintaining this distribution.



Sodium Potassium- ATPase pumps. Image created at BYU-Idaho by MG 2013
Three Na^+ ions are moved out of the cell in exchange for two K^+ ions with the aid of ATP.

Secondary Active Transport

Like primary active transport, secondary active transport also moves solutes against their concentration gradients. However, with secondary active transport, ATP is not directly involved in the pumping of the solute. Instead, this process

uses the energy stored in concentration gradients to move the solute. Since sodium is always in higher concentration outside of the cell (due to primary active transport), the sodium gradient is often used to power secondary active transport. In this process, the carrier protein has a binding site for the solute to be transported against its concentration gradient and a binding site for sodium. Once both solutes have bound, sodium moves down its concentration gradient into the cell, much like what happens with carrier-mediated diffusion and in the process provides the energy required for the other solute to be transported into the cell (**symport**) or out of the cell (**antiport**), against its concentration gradient. A number of organic molecules are transported across membranes by this process, such as glucose and amino acids. ATP energy is required to generate the sodium concentration gradient but is not directly involved in moving the desired solute across the membrane. It is the dissipation of this sodium gradient that provides the energy required for *secondary active transport.*

Bulk Transport

To this point, we have been talking about the movement of relatively small solutes across the cell membranes (i.e. ions and small organic molecules). There are instances, however, when it is necessary to move much larger materials across the membrane, like when a macrophage engulfs a bacterium or when larger amounts of a given material are released from a cell, such as the release of a hormone. These processes also require ATP and are, therefore, examples of active transport, but they move materials in a very different way.

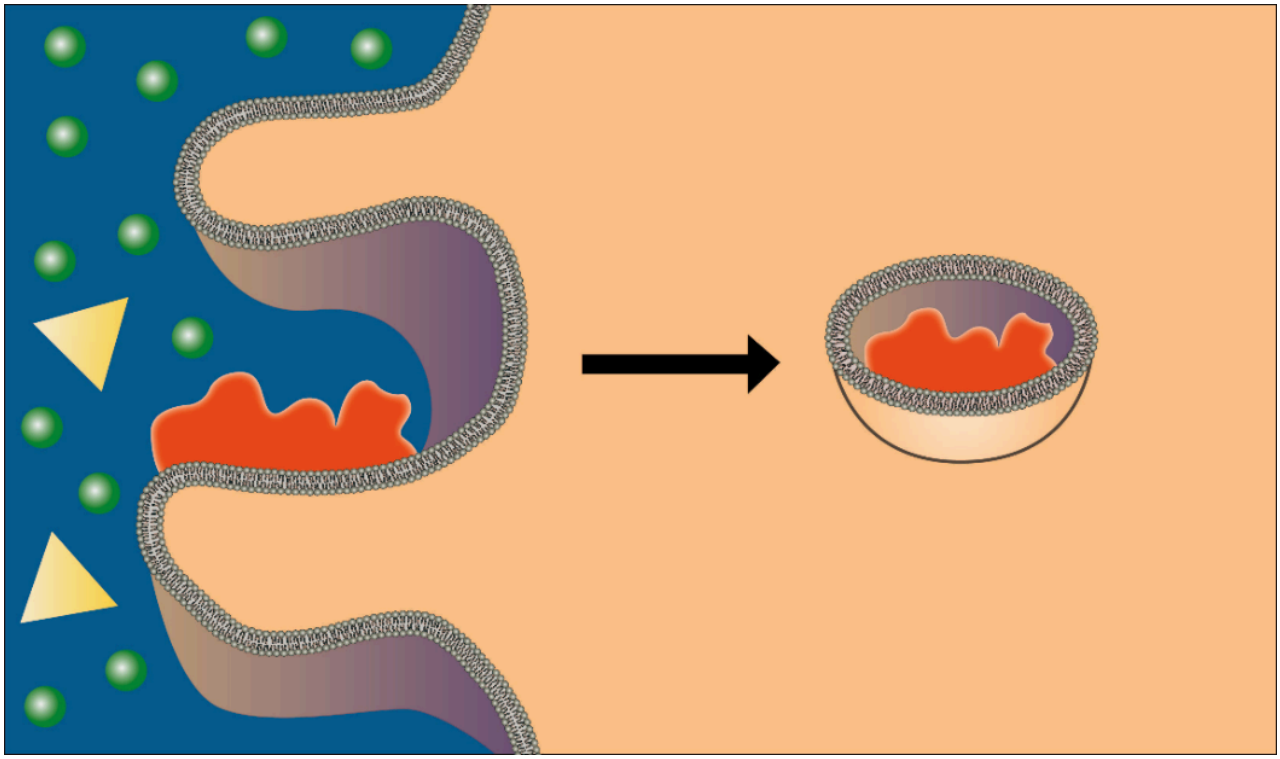
Endocytosis

Endocytosis is the bulk transport of materials into the cell. There are several types of endocytosis, and we will briefly explore each one. First, let's discuss **phagocytosis** (see figure below), which means *cell eating*. Only a limited number of cells are capable of phagocytosis, specifically cells of the immune system. In this process, the cell sends extensions of its plasma membrane, called ***pseudopodia***, out and around the particle to be phagocytized. As these pseudopodia surround the particle, they eventually fuse, creating a vesicle containing the particle. This **phagosome** can then unite with a lysosome inside the cell, and the engulfed material can be digested for use within the cell. Watch an amoeba phagocytize a paramecium in this clip:

https://books.byui.edu/-SLP (Transcription Available)

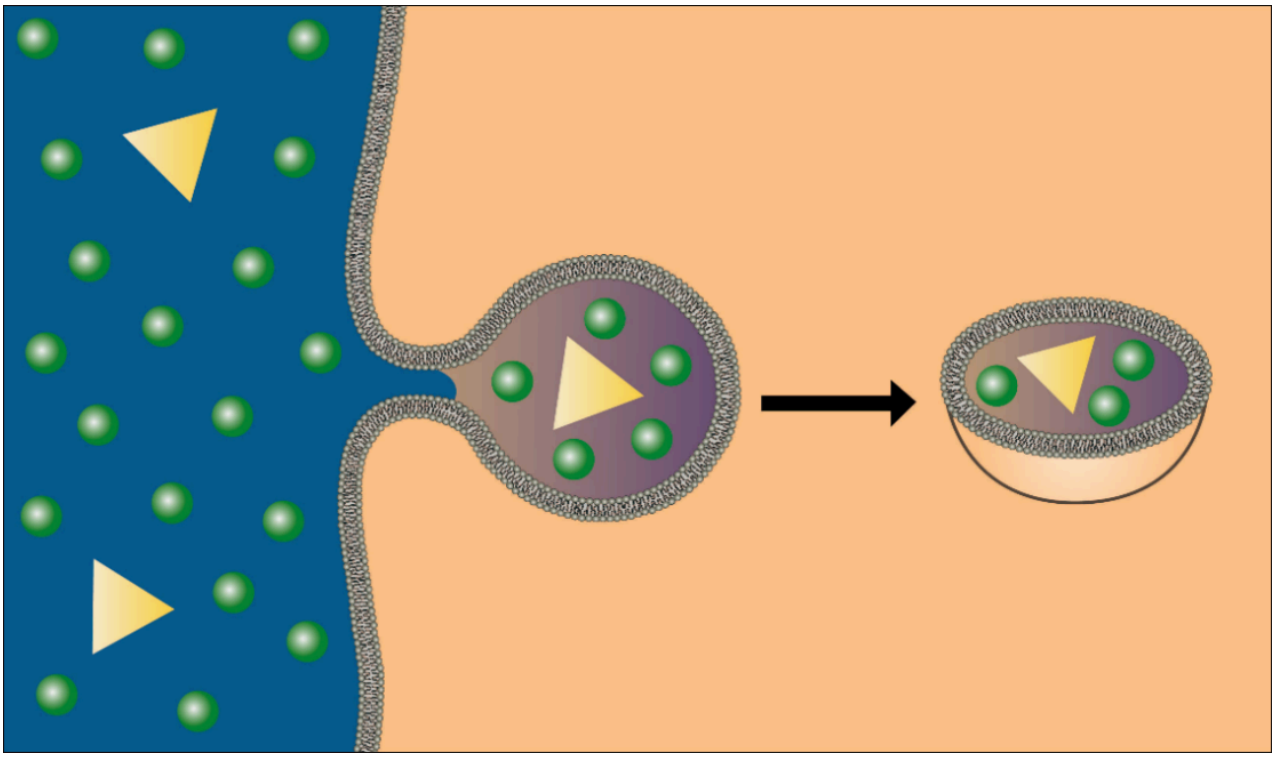
In this next clip, watch a neutrophil, one of our white blood cells, chase down and phagocytize a bacterium (A humerus musical approach to viewing this).

https://books.byui.edu/-qxyj



Phagocytosis. Image created by BYU-Idaho student, Hannah Crowder, 2013.
In phagocytosis (shown above), the cell membrane forms processes that surround and engulf a particle to be brought into the cell.

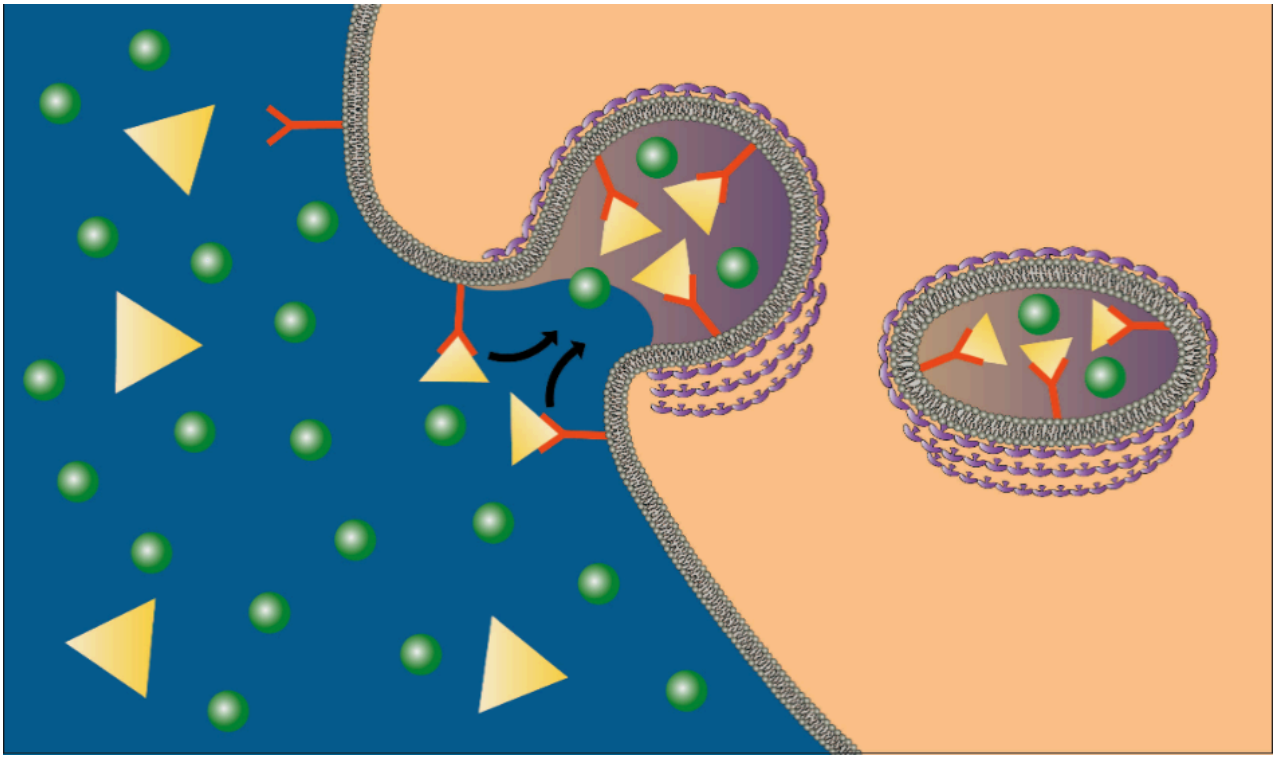
A second type of endocytosis is **pinocytosis**, which means *cell drinking*. In this process, rather than send out pseudopodia, the cell membrane simply invaginates (forms a pocket) and engulfs anything in the fluid that is taken into the cell (see figure below). Unlike phagocytosis, pinocytosis occurs in most cells of the body. The cells are not interested in the water in the vesicles but any solutes that might be brought in. As you can imagine, this is not a very efficient way of bringing materials into the cell because it is nonspecific and brings whatever is in the fluid into the cell. It provides cells with a nonselective mechanism for sampling the extracellular environment. It is prominent in cells involved in moving large amounts of material across the membrane, like cells of the intestines and the kidneys.



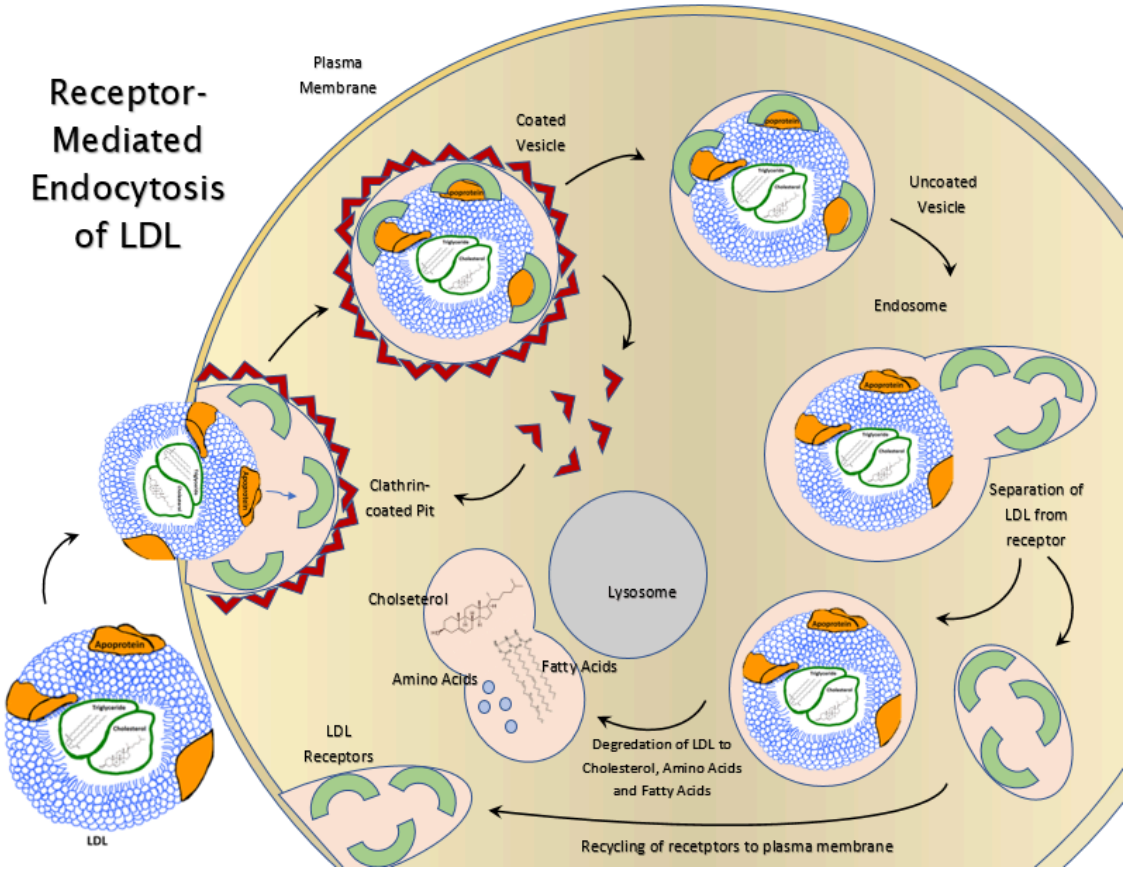
Pinocytosis. *Image created by BYU-Idaho student, Hannah Crowder, 2013.*

In pinocytosis, the membrane forms an invagination (pocket) that pinches off, bringing into the cell the fluid in the pocket along with any solutes in the fluid.

A much more efficient mechanism for bringing specific solutes into the cell is **receptor-mediated endocytosis**. As the name implies, this mechanism employs specific receptors that bind to specific compounds (**ligands**) within the extracellular space. Once the specific ligand binds with its receptor, the resulting complex migrates to a specific area of the membrane called a clathrin-coated pit. The clathrin protein is activated by the bound receptor which initiates endocytosis in a process similar to pinocytosis (see figure below). The advantage of receptor-mediated endocytosis is that it can engulf large amounts of a specific solute. The following two images demonstrate how this process occurs. The first is a general mechanism for receptor-mediated endocytosis, and the second shows how a specific molecule, cholesterol, is brought into the cell by this process.



Receptor-Mediated Endocytosis. *Image created by BYU-Idaho student, Hannah Crowder, 2013.*
In receptor-mediated endocytosis, ligands bind to specific receptors, which then migrate to a clathrin-coated pit. The contents are then brought into the cell by a process similar to pinocytosis. Below a LDL particle is taken into the cell through endocytosis to retrieve cholesterol molecules.



Receptor-Mediated Endocytosis of LDL. *Image created for BYU-Idaho by T. Orton, 2017*

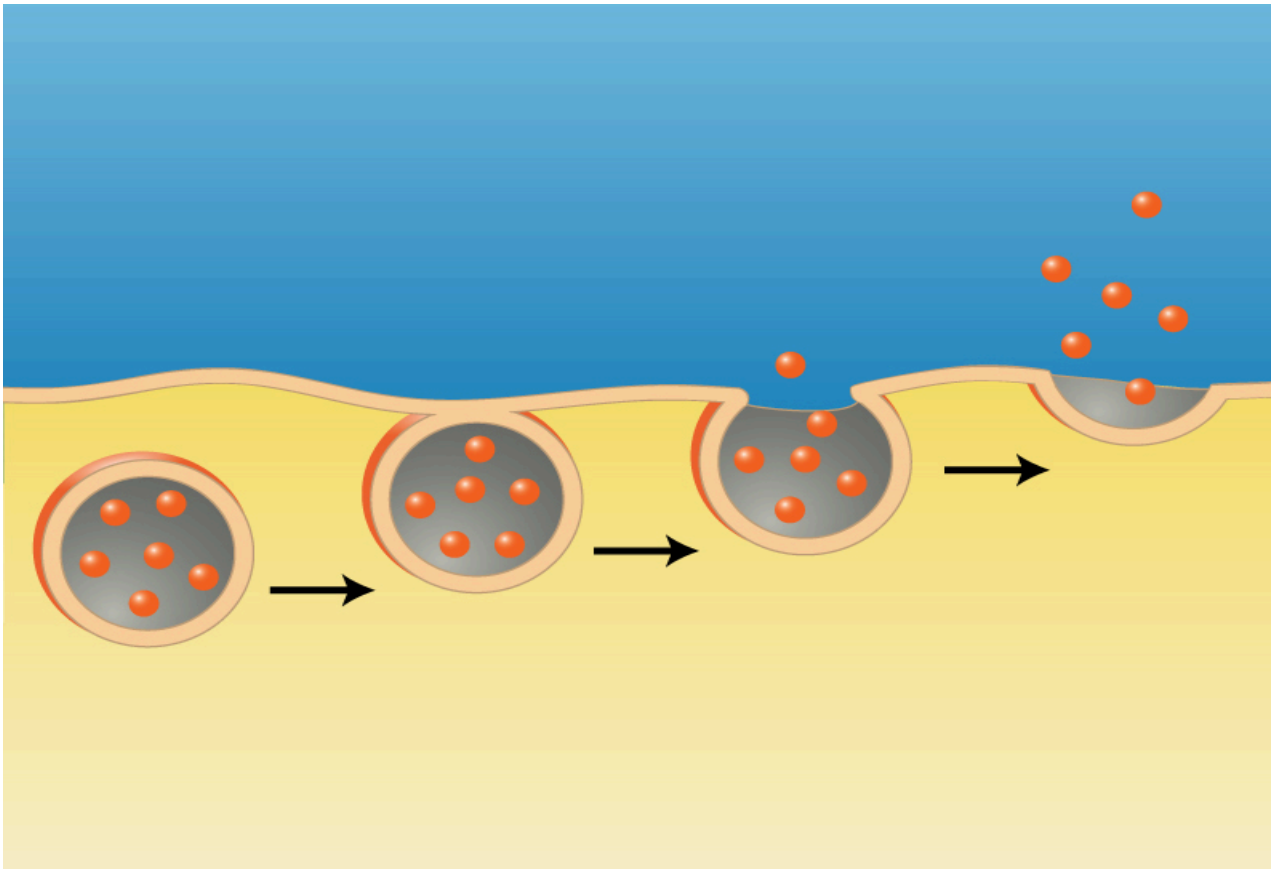
Exocytosis

Thus far, we have been discussing bulk transport, bringing material into the cell. There is also a need to export material from the cell into the extracellular fluid. This process is called **exocytosis**. Exocytosis is the process by which the beta cells of the pancreatic islets secrete insulin into the extracellular fluids. The mechanism is essentially the reverse of endocytosis. **Secretory vesicles** filled with the material to be released migrate to the plasma membrane where the membrane of the vesicle fuses with and actually becomes a part of the plasma membrane (see figure below). The material that was in the vesicle suddenly finds itself outside of the cell, and any integral protein within the vesicle membrane now becomes a protein expressed on the cell membrane (See example below of GLUT4). While this is a complex process, the usual signal that initiates exocytosis is the entry of calcium ions into the cell which bind to specific proteins (e.g. SNARE proteins) that initiate this process. Since calcium concentration is higher outside the cell, it is common that activation of gated calcium channels in the membrane precede exocytosis.

The links below show how endocytosis and exocytosis work.

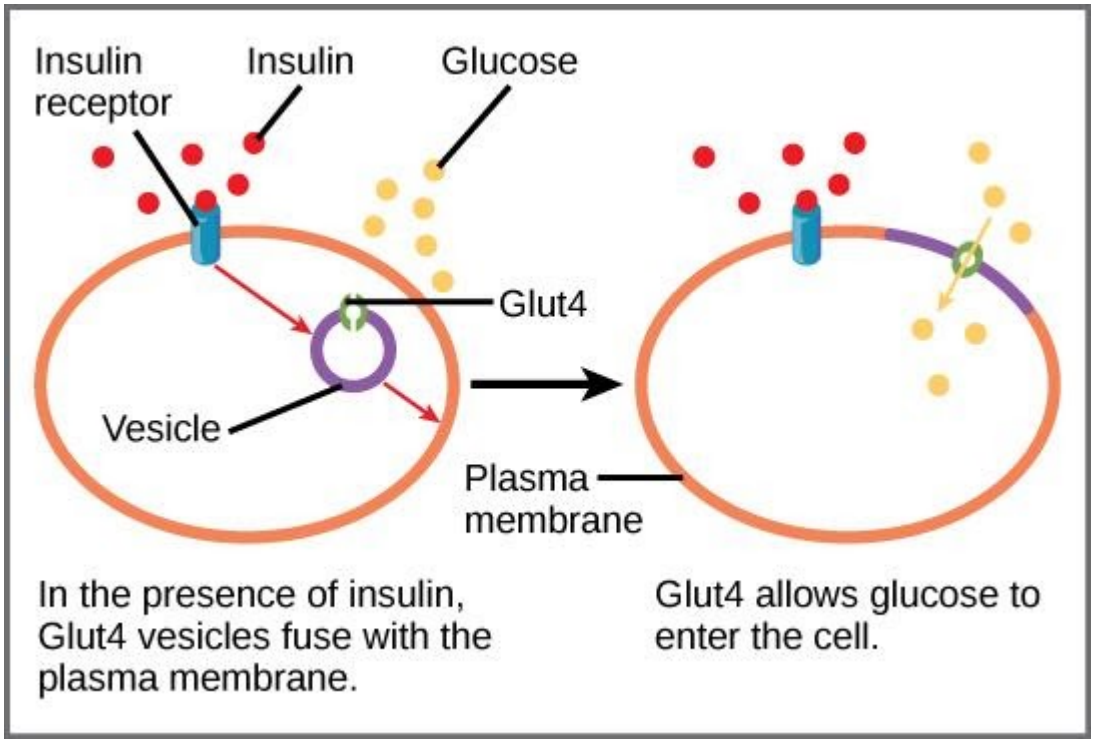
https://books.byui.edu/-yFo (Transcription Available)

https://books.byui.edu/-kANn



Exocytosis. *Image created by BYU-Idaho student, Hannah Crowder, 2013.*
In exocytosis, secretory vesicles migrate to the cell membrane where the vesicular membranes fuse with the plasma membrane, releasing the vesicles' contents into the extracellular fluid.

In skeletal muscle, GLUT4 proteins (glucose transporters) are found in intracellular vesicular membranes. When insulin binds to its cell surface receptor, exocytosis is initiated to allow GLUT4 expression on the plasma membrane and subsequent glucose uptake.



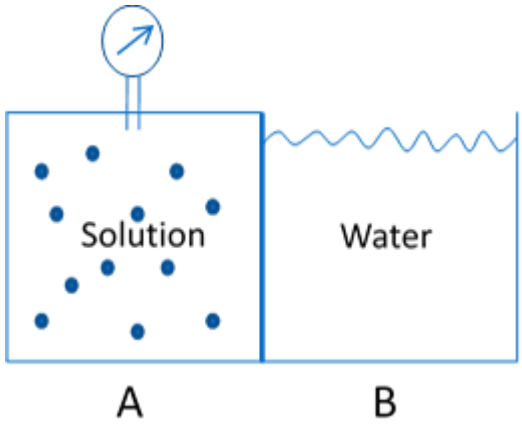
GLUT 4 Carrier Protein. *By CNX OpenStax [CC BY 4.0 (https://books.byui.edu/-CjnG], via Wikimedia Commons*

5.2.4

Osmosis

A special type of passive transport is the movement of water across a membrane, or **osmosis**. By definition, osmosis is the diffusion of water through a **selectively permeable** membrane from an area of high water potential (low solute concentration) to an area of low water potential (high solute concentration). Therefore, for osmosis to occur, the membrane must be permeable to water and the concentration of the solute must be different on the two sides of the membrane. Water will move from the side with lower solute concentration to the side with higher solute concentration until the concentrations are equal or until some external force prevents further movement of water. This is a passive facilitated process, in that no energy is required for the movement of water through specialized channel proteins known as ***aquaporins***. All cells of the body express aquaporin proteins which enable osmosis to occur across cellular membranes.

Let's start to apply this information. In an artificial system such as the one depicted in the figure below, water will attempt to move from chamber B to chamber A across the water-permeable membrane separating the two chambers. This will occur because chamber A has more solute dissolved in it than in chamber B (no solute). Since chamber A is a rigid chamber, pressure will develop. The pressure that is just sufficient to prevent water from moving across the membrane is referred to as **osmotic pressure**. Osmotic pressure can also be defined as the force exerted by the process of osmosis (diffusion of water).



Osmotic Pressure. Minimum pressure needed to be applied to a solution to prevent the inward flow of water across a semipermeable membrane from chamber B to chamber A. *Image by BYU-Idaho student, Hannah Crowder, 2013.*

Similar to this system, water will move into or out of cells depending on the solute concentration of the extracellular fluids and the intracellular fluids. If the solute concentration in the extracellular fluid is lower than the solute concentration inside the cell, water moves into the cell and the cell will swell and potentially burst! If the opposite is true (extracellular solute concentration is greater than intracellular), then water will leave the cell causing it to shrivel, or crenate. This phenomenon is due to osmosis and it is critical that you understand how to predict this activity since the cell shape directly affects its functionality.

Earlier this semester, we discussed how total solute concentration, or the total number of moles of particles dissolved per liter of solution is expressed as **osmolarity**. This is different than molarity. Molarity represents the number of moles of a specific compound per liter of solution. Why do we have these different ways of expressing concentration? In the cellular environment, the extracellular and intracellular fluid are complex mixtures of numerous dissolved solutes, all of which contribute to the osmotic pressure driving osmosis. To understand osmosis, we need to consider the osmolarity of the extra- and intracellular compartments.

For example, when 1 mole of NaCl is dissolved in 1 liter of water, it is a 1 molar (M) NaCl solution. However, NaCl is an ionic compound that dissociates (breaks apart) into 1 mole of Na^+ and 1 mole of Cl^- ions when dissolved in water. Thus, there are now twice as many particles (Na^+ and Cl^- ions) prior to dissolving it and we call it a 2 osmolar (OsM) solution (simply add the number of moles of dissolved ions). Glucose is different. Glucose doesn't dissociate in water because the atoms are covalently bonded. Therefore, a 1 M solution of glucose (1 mole of glucose in 1 liter of water) will be a 1 OsM solution. What if you had a 1 liter solution that contained 1 mole of NaCl and 1 mole of glucose? The molar concentration would be 1 M NaCl and 1 M glucose, but what would the osmolarity be? Simply add the moles of dissolved particles: 1 mole Na^+ + 1 mole Cl^- + 1 mole glucose = 3 moles of particles per 1 liter. This solution would be a 3 OsM solution.

In physiological solutions like the intracellular or extracellular fluid, there are many solutes dissolved including ions, proteins, nutrients, and waste products. All of these dissolved components contribute to the osmolarity of the fluid. The normal osmolarity of body fluids is 285–295 mOsM (for simplicity we often round this number to 300 mOsM…you need to remember this number). We place the small m, which stands for milli or one-thousandth, in front of OsM because we are dealing with very small amounts, 1000 times less than 1 osmole/liter. Osmolarity is often used to compare two solutions using the following terms: *isosmotic*, meaning two solutions have the same osmolarity; *hyperosmotic*, meaning one solution is more concentrated than the other; and *hypoosmotic*, meaning one solution is less concentrated than the other. In the previous figure illustrating osmotic pressure, you could say, “Solution A is hyperosmotic to solution B” because solution A has greater osmolarity than solution B. Always use the term osmolarity to describe solutions that have not been administered to cells (no cells are present).

In more complex physiological environments, the osmolarity of the extracellular environment and the process of osmosis (water diffusion) across cellular membranes play a significant role on the function of cells. When water moves out of a cell, the cell shrinks; likewise, when water moves into a cell, the cell swells. These changes in cell shape affect membrane and cell function and can even cause the cell to die. The term **tonicity** is used to describe these osmotic effects on cells based on the solution or extracellular environment the cells are in. Remember that water diffuses to areas of greater solute and like osmolarity, the prefixes iso-, hypo-, and hyper- are used to describe tonicity. If red blood cells (RBCs) are placed into a 300 mOsM solution containing Na^+ and Cl^- ions which cannot diffuse into the cells, then there will be no net movement of water into or out of the RBCs because the intracellular osmolarity is also 300 mOsM. We would classify this solution as ‘isotonic’ (note that it is not isosmotic because cells are present). In the healthcare setting, this isotonic solution is called physiologic saline, or 0.9% NaCl, or 0.9% saline. The 0.9% NaCl solution is about 300 mOsM. If RBCs were placed into a different NaCl solution (no diffusion of ions into the cell) with a 600 mOsM osmolarity (or 1.8% NaCl solution), water would diffuse out of the RBCs to the area of higher solute, causing them the cells crenate. This solution would be considered ‘hypertonic.’ If RBCs are placed into distilled water (i.e. no solute), then water will diffuse into the RBCs (following solute) causing them to swell and potentially burst (called ‘hemolysis’). This type of environment is considered ‘hypotonic.’ This may seem simple, but it can get quite complex especially if the membrane is permeable to the solute in the extracellular fluid.

Understanding membrane permeability, osmolarity, and osmosis will enable you to properly predict the resulting tonicity of various solutions when administered to a person intravenously and the resulting effects on the tissue. Always use the term tonicity to describe the effects of a solution in the presence of cells. Let’s consider some examples involving RBCs to better understand tonicity:

A patient comes to the clinic having overdosed on insulin and is severely hypoglycemic (really low blood glucose). A 5% glucose (sometimes called dextrose) solution (~300mOsM) is prepared and administered to the patient. The patient’s

blood glucose increases but then the patients RBCs begin to swell and lyse (burst). What is happening?

If the cells are swelling, that means water is diffusing into the cells, so the RBCs must be in a hypotonic environment. How can this be if the solution of glucose administered (5%) was isosmotic with the inside of the cell (both ~300mOsM)? Unlike NaCl, RBCs express GLUT proteins that allow glucose to diffuse into the cells. Once in the cell, the cell metabolizes the glucose for energy. Then more glucose enters the cell and is further metabolized. Eventually, all of the 5% glucose solution is metabolized by the cell and nothing is left but the water. That is why the cells, upon reaching equilibrium began to swell. As the glucose was diffusing into the cells, the extracellular environment was becoming less concentrated and more hypotonic, causing water to diffuse into the cell. This is why glucose is typically administered in a 0.9% NaCl solution. The glucose-saline solution is hyperosmotic to the intracellular environment (>300 mOsM), but after administration to the patient, the glucose diffuses into the cell, gets metabolized, and the remaining NaCl solution maintains an isotonic environment for the cells.

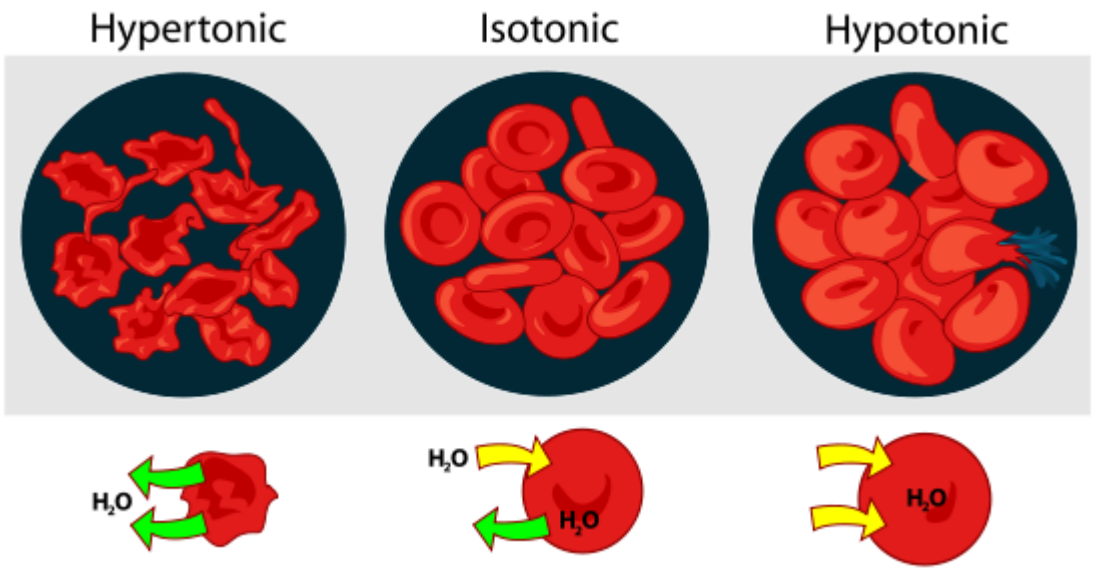
Here are three rules to help you better understand tonicity:

1. In order for ions or polar (hydrophilic) molecules to cross the cell membrane, there needs to be channel or carrier proteins expressed on the cell membrane. Then, solute will diffuse across the membrane until equal concentrations are reached on both sides of the membrane (equilibrium).
2. If solute cannot diffuse across the membrane, then water will diffuse into, or out of the cell in attempt to balance the difference in osmolarities between the intra- and extracellular environments.
3. Based on the direction of osmosis (into or out of cell) we can classify the extracellular environment as iso-, hypo-, or hypertonic.

Consider what might happen with a membrane-permeable solute that moves into the cell but does not disappear (e.g. urea). If we start with an isosmotic solution of urea (~300 mOsM) and then place a cell into this solution, the urea, a membrane-permeable solute, will move down its gradient into the cell. We say down the gradient because initially, the concentration of urea outside the cell is much greater than the concentration inside the cell. Urea will move into the cell until the urea concentration inside the cell equals the urea concentrations outside the cell. In other words, 150 milliosmoles of urea will eventually diffuse into the cells, leaving 150 milliosmoles in the extracellular solution. As this diffusion of urea is taking place, what is happening to the intracellular osmolarity? Theoretically, the intracellular osmolarity should reach 450 mOsM (300 mOsM + 150 mOsM urea) while the extracellular solution should approach only 150 mOsM urea. What is water going to do in this situation? It will diffuse into the cell in attempt to balance out the unequal osmolarities between the intra- and extracellular fluids to reach equilibrium. The cells will swell, and this solution will be classified as hypotonic. A good rule to remember is that any isosmotic solution of a penetrating or permeable solute will act as a hypotonic solution to the cell.

Here is another way to think of osmolarity and tonicity. Osmolarity can be used to compare the concentration of solutes in two solutions. It can also be used to compare the osmolarity of a solution with that of the cell before equilibrium is achieved, or before the solution is administered to cells. Tonicity is used to describe what effect the solution has on the cell. Osmolarity doesn't take into account the permeability of the solutes, while tonicity is dependent upon the concentration of the *nonpermeable* solutes.

The figure below shows what happens to red blood cells when they are placed into hypertonic, isotonic, or hypotonic solutions.



Osmotic Pressure on Blood Cells Diagram. *Title: File: Osmotic pressure on blood cells diagram.svg; Author: LadyofHats;*

Site: https://commons.wikimedia.org/wiki/File:Osmotic_pressure_on_blood_cells_diagra

License: Public Domain

The link below shows what happens to a wilted plant when it is placed into a hypotonic solution.

https://books.byui.edu/-vip

To check understanding, complete the table below by filling in the missing column items with regard to osmolarity and tonicity. Use the terms *iso*, *hypo*, and *hyper* to complete the table.

SOLUTION	OSMOLARITY (No cells, but compared to normal intracellular osmolarity 300mOsM)	TONICITY (Solution is administered to cells and equilibrium is being reached)
0.9 % saline		
5% dextrose		
5% dextrose + 0.9% saline		
0.45% saline		
5% dextrose + 0.45% saline		

Here are the answers for the table above. Be sure you understand why the answers are what they are.

SOLUTION	OSMOLOARITY	TONICITY
0.9% saline	Isosmotic	Isotonic
5% dextrose	Isosmotic	Hypotonic
5% dextrose + 0.9% saline	Hyperosmotic	Isotonic
0.45% saline	Hypoosmotic	Hypotonic
5% dextrose + 0.45% saline	Hyperosmotic	Hypotonic

****Note:** Other texts, even hospitals on occasion, tend to use less rigorous definitions of tonicity. For example, definitions are loosely given to define all hyperosmolar solutions as hypertonic. This is based on the observation that water can cross the membrane faster than the permeable solute can cross. It may also be based on the incorrect assumption that tonicity and osmolarity are the same thing. Having a more accurate understanding of osmolarity and tonicity will not only benefit you in this course, but it will also enable you to better understand the relationship between chemistry and cell biology.

5.3

INTRODUCTION TO ELECTROPHYSIOLOGY

One of the most difficult concepts you will encounter on your journey to becoming a budding student of Anatomy and Physiology is **electrophysiology**. Electro—what? Exactly! However, we are confident that after a few light pages of reading, you will be able to use electrophysiological principles in your everyday conversations. Why is this subject so important? For starters, consider this quote from President John Taylor: "...I could show you upon scientific principles that man himself is a self-registering machine, his eyes, his ears, his nose, the touch, the taste, and all the various senses of the body, are so many media whereby man lays up for himself a record..." (Pres. John Taylor, *Journal of Discourses*, 26:32.). In more scientific terms, the body is able to take external stimuli (i.e. light, sound, smell, taste), using various sensory organs of the body, and convert these stimuli into unique perceptions in the brain. How the body converts the various types of stimuli from outside sources to inside signals is a major emphasis of electrophysiology. In addition, consider the magic that occurs when your brain decides to put your hand in your pocket and, using mechanoreceptors on your fingers, pull out the exact item you were seeking, without even looking at it! Well, with any luck, after you are done with this section, you should be able to explain the "magic" behind these different phenomena.

5.3.1

Ions and Cell Membranes

Review of Ions

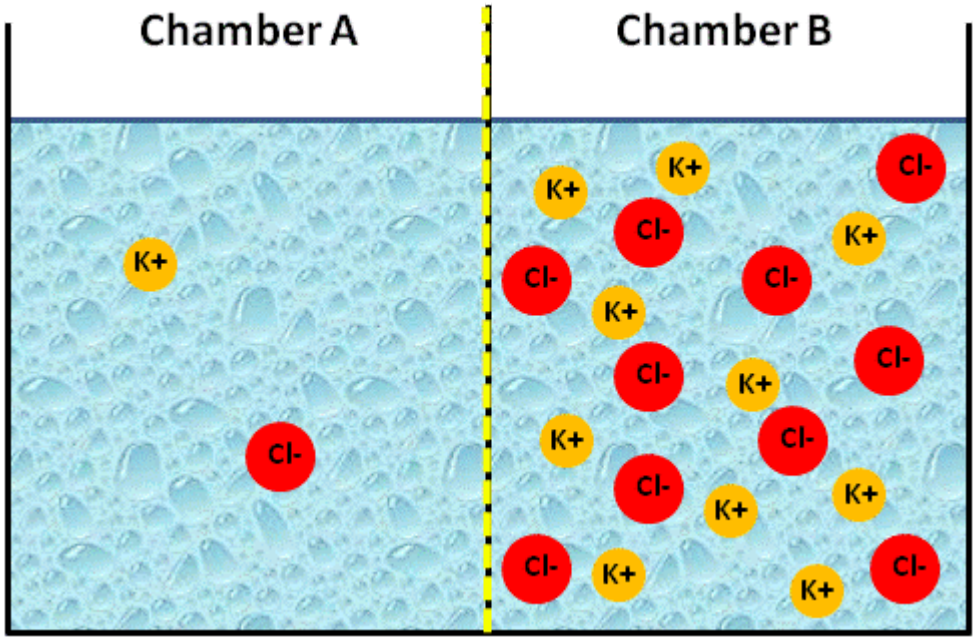
Electrophysiology, by definition, is the study of the electrical properties of biological cells. It involves measurements of electric current and electrical activity of not only neurons, but virtually all cells of the body. When we talk about the electrical properties of cells, we are not talking about electrons moving through wires like you may suspect. Rather, electrophysiology deals with ions (not electrons) and their movements across the cell membrane. You have learned that atoms consist of equal numbers of positively charged protons and negatively charged electrons, unless they give up or accept electrons, which would make them an ion. When we take an ionic compound like NaCl and place it into water, the two ions dissociate and become free floating Na^+ (positive charge, cation) and Cl^- (negative charge, anion) ions. These ions (Na^+ and Cl^-) along with Ca^{++}, H^+, and K^+ are very important in the human body.

Separating Charges

Opposite charges attract, similar to how Na^+ and Cl^- ions are attracted to each other to form an ionic compound. If we can somehow separate opposite charges from one another and then keep them separated, but maintain the possibility of reuniting, we can generate quite a bit of energy. Consider what happens when you rub a balloon on your head. When we add work (energy) by rubbing the balloon on the head, we cause a separation of electrical charge. In other words, we cause the electrons from your hair proteins to move from your hair to the balloon surface. The balloon surface gains an attractive force called *static electricity*. When we put the balloon close to other people's hair, their hair shafts will become attracted to the negative charges on the balloon. Two important things come out of this analogy; first, opposites attract, and second, it takes energy, or work, to separate charges. In the body, the cell membrane has the ability to separate charges (i.e. ions) by performing work. This separation occurs through the action of an ATP driven pump called the *sodium/potassium ATPase pump*. Every cell of your body expresses this protein for the specific purpose to separate Na^+ and K^+ ions across the membrane.

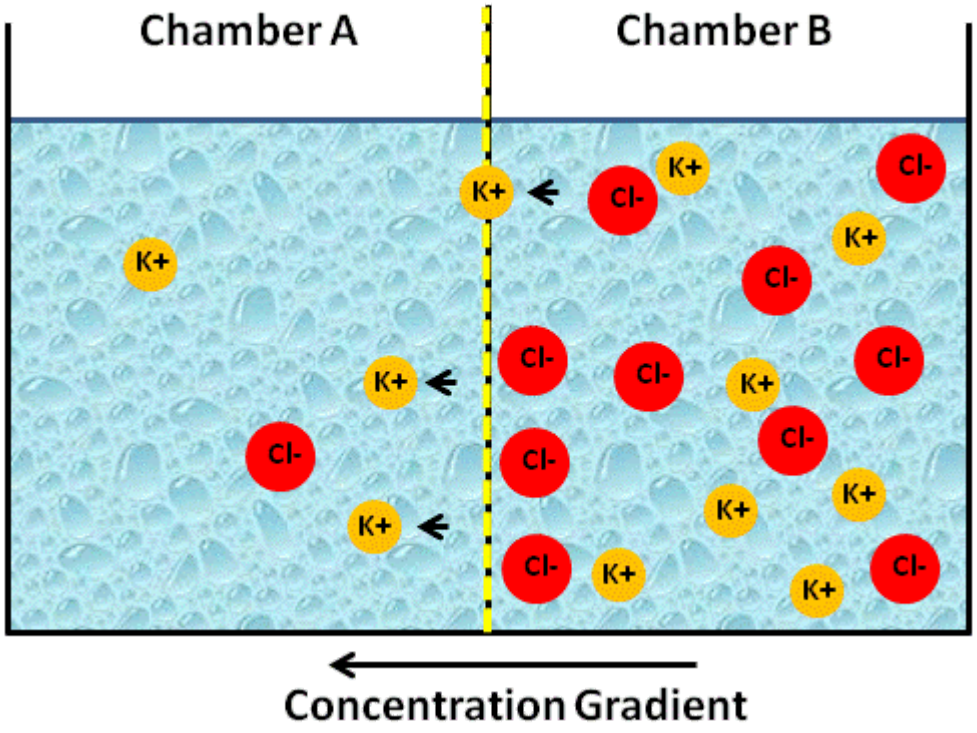
Two Chamber Analogy

The figure below represents an artificial system of two chambers (A and B) separated by a semipermeable membrane. Semipermeable means that only certain substances can diffuse across it. We will assume that the membrane is permeable to cations, but not to anions. If we placed two different concentrations of a KCl solution into each chamber, say 0.1 M solution of KCl in chamber A and a 1.0 M solution of KCl in chamber B, what would happen (See figure below)?



Two Chamber Analogy 1: Initial Ion Arrangement. *Image Created at BYU-Idaho by JH, 2013.*

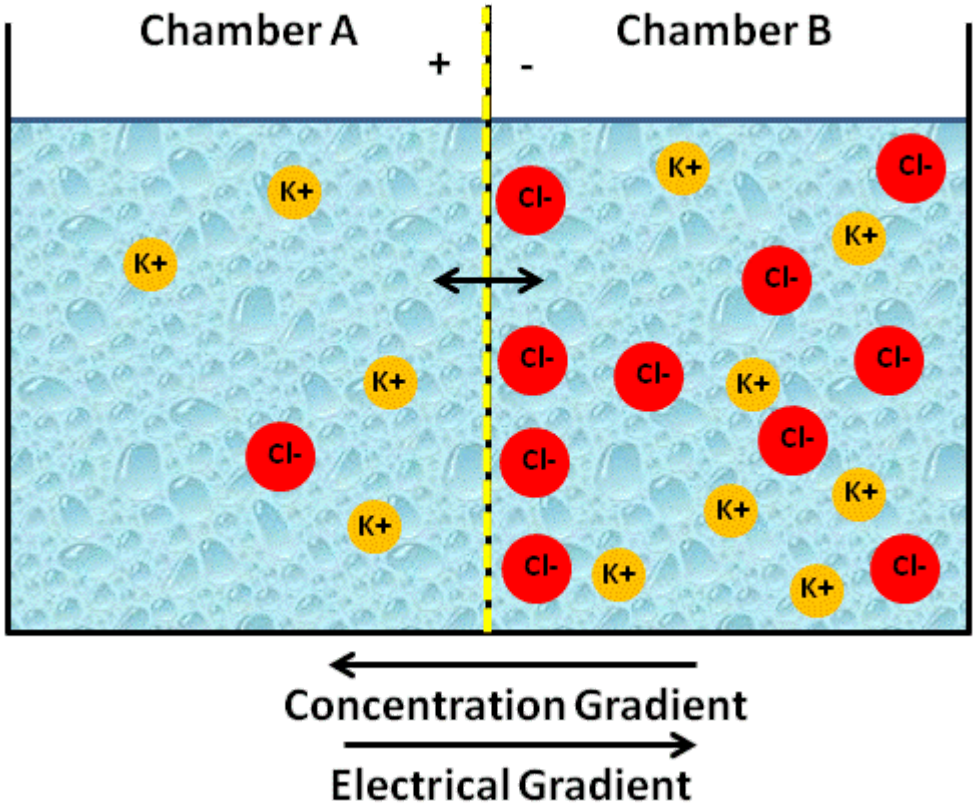
Since the membrane is only permeable to cations, K^+ will begin to diffuse down its concentration gradient from chamber B (high K^+ concentration) to chamber A (low K^+ concentration (see figure below).



Two Chamber Analogy 2: K+ moving down concentration gradient from Chamber B to A. *Image Created at BYU-I by JH, 2013.*

When will the diffusion of K^+ stop? You might be tempted to say, "When the concentration of K^+ in chamber A is equal to the concentration of K^+ in chamber B." In the previous section, you would have been correct! But in this section, we are

dealing with charged solutes (ions). When a single ion diffuses across a membrane, it creates an electrical gradient. Similar to a concentration or chemical gradient, an electrical gradient is when you have an area of higher charge than another. Look back at the figures above., Initially, no charge differences existed between the two chambers; both sides had the same number of positive and negative ions respectively. However, since the membrane is not permeable to Cl^-, and K^+ started to diffuse into side A going down its concentration gradient, there is now a difference in charges between the two chambers. Side A now has more positive ions (K^+) than negative ions, making it positively charged; and side B now has more negatively charged ions (Cl^-) than positive, making it negatively charged (see figure below). With chamber A being positively charged and chamber B being negatively charged, we have now created an electrical gradient. For K^+, a cation that is attracted to negatively charged particles (i.e. the Cl^- ions), its electrical gradient actually counters its concentration gradient, going from side A to B. This means that it becomes more difficult for K^+ to diffuse down its concentration gradient into side A because of the electrical force (positive charge) it generates that opposes more K^+ from diffusing across the membrane. Remember opposites attract, so the Na^+ is more attracted to the chamber with more Cl^- ions.



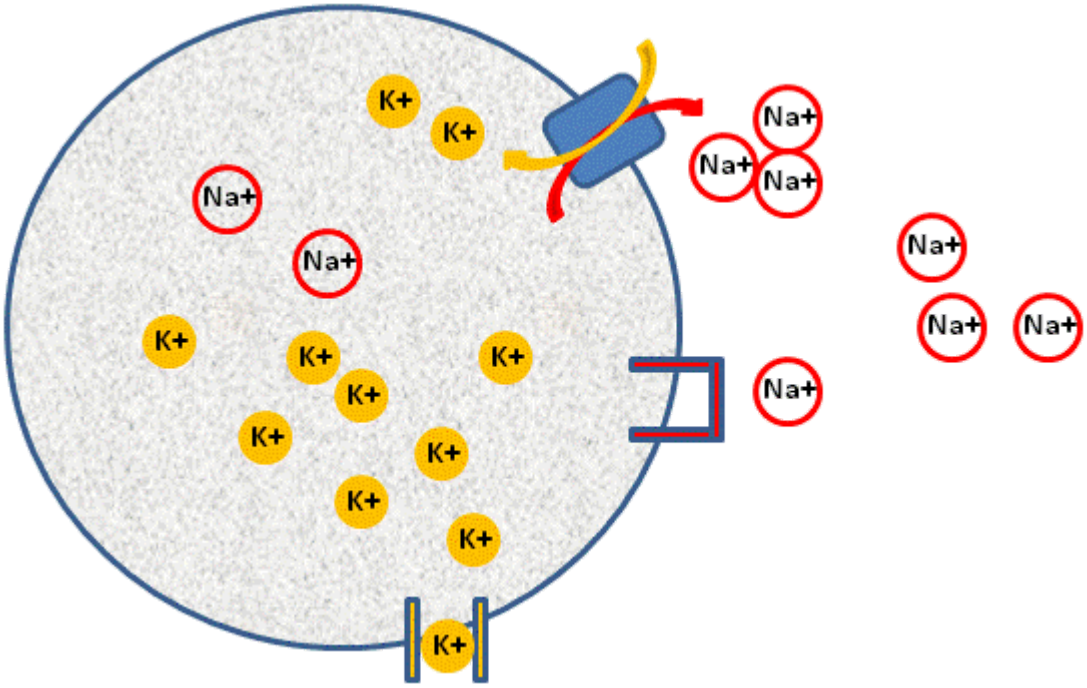
Two Chamber Analogy 3: Electrical gradient pulls K+ ion back to chamber B preventing it from reaching equilibrium.
Image Created at BYU-I by JH, 2013.

The net flow of K^+ will stop when the force of the electrical gradient (i.e. the attraction of the Na^+ ions to the Cl^- ions) equals the force of the concentration gradient. We call this a state of equilibrium, even though the numbers of K^+ on side B will still be higher than side A. When the two forces, the electrical gradient and the chemical gradient, equal each other (equilibrium) and the net movement is zero, we call the state resting and can refer to the two combined gradients as the **electrochemical gradient**. We have now successfully created a state of charge separation, something we call an electrical potential, or voltage because side B is now negatively charged compared to side A.

5.3.2

Membrane Potentials

In the cell, we use the same principles as the two-chamber analogy, except that we initially use energy and work, to separate the ions from each other (think of our balloon and hair experiment described above). As mentioned above, this work is primarily accomplished by the Na^+/K^+ ATPase pump. This pump moves three Na^+ ions out of the cell and two K^+ ions into the cell for every ATP (energy) molecule hydrolyzed (see figure below).



Resting Membrane Potential. *Image created at BYU-Idaho by JH, 2013.*

Diagram of a cell establishing a resting membrane potential. Shown are representations of a K^+ channel that allows for the movement of K^+ out of the cell, a closed Na^+ channel that prevents movement of Na^+ into the cell, and a Na^+/K^+ ATPase pump that moves three Na^+ ions out of the cell and two K^+ ions into the cell for every ATP molecule hydrolyzed. ATP hydrolysis is not shown.

This pump is found in every cell in the body. It creates a concentration gradient for both K^+ and Na^+. It causes a much higher concentration of K^+ to exist inside the cell and a much higher concentration of Na^+ to exist outside the cell. This results in a very large concentration gradient for K^+ to leave the cell and a very large concentration gradient for Na^+ to come into the cell. While the cell membrane expresses Na^+ channels, they remain

closed under normal ‘resting’ conditions and prevent Na^+ from diffusing into the cell. The cell membrane also expresses K^+ leak channels (no gate to regulate K^+ permeability) that allow K^+ ions to leave the cell. As a result of these differences, the membrane is 50–100 times more permeable to K^+ than Na^+ under resting conditions.

Resting Membrane Potential

Remember that the Na^+/K^+ ATPase pump transports 3 Na^+ ions out and only 2 K^+ ions into the cell every cycle. This means that over time the outside of the cell will become more positively charged because more positive ions are being pumped out of the cell than into the cell. Additionally, due to the K^+ leak channels expressed on the cell membrane, K^+ will start to diffuse out of the cell going down its concentration gradient. As K^+ leaves the cell, it will leave behind negative charges, just like what happened when the K^+ diffused to side A in our two-chamber analogy illustrated above. Inside the cell however, most of the negative charge comes from negatively charged proteins. The result of this K^+ diffusion is the same as the analogy: the inside of the cell will start to become negative with respect to outside of the membrane. As more K^+ diffuses out of the cell, the inside of the cell becomes more negative. When the membrane reaches a state of equilibrium (i.e. the chemical gradient driving K^+ out of the cell equals the electrical gradient forcing K^+ to remain in the cell), the difference in charge between the intra- and extracellular environments is called the **resting membrane potential (or voltage)**. We can actually measure the negative charge inside the cell and express this measurement in millivolts (mv). It is critical that you understand this concept. Remember that the resting membrane potential is dependent on the 3 things described above: 1) The work of the Na^+/K^+ ATPase pump; 2) the diffusion of K^+ out of the cell via K^+ leak channels; and 3) the negatively charged proteins that remain in the cell.

Membrane Potentials and Excitable Tissues

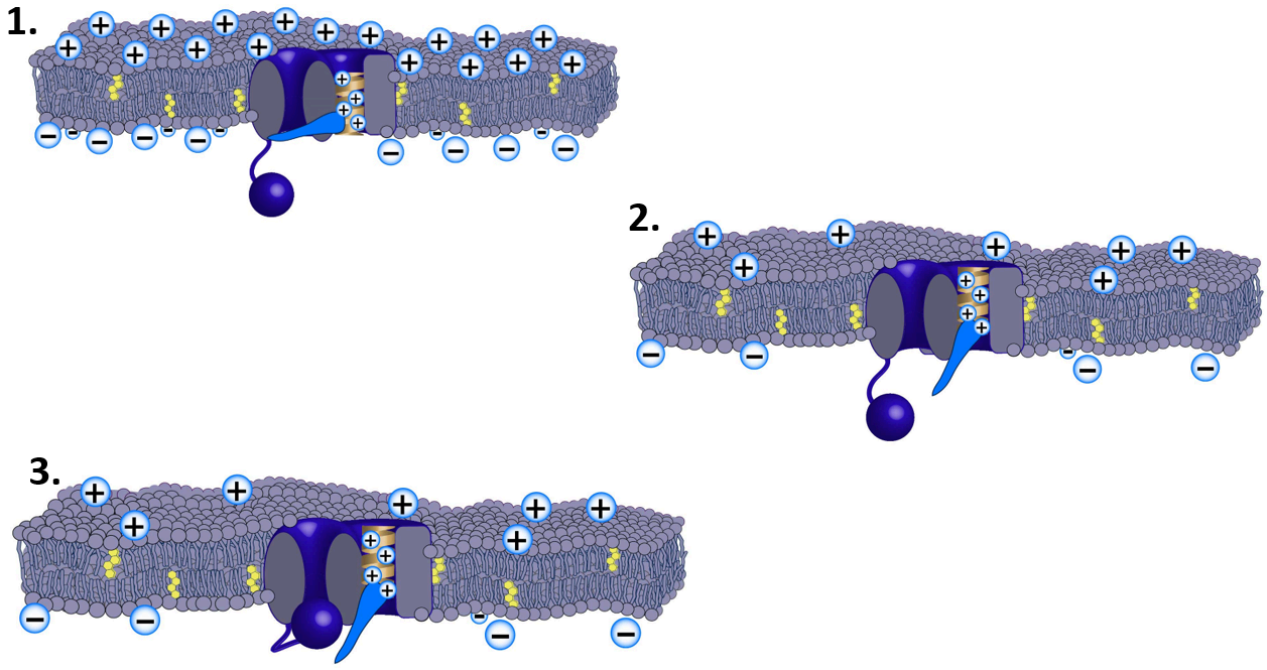
So what? Who cares if the inside of cells are negative? You may know that a battery, like the one that powers your phone, is nothing more than a separation of electrons. When you connect the ends of a battery with a wire, those electrons flow through them and can do work, like light your screen. How do you measure how strong a battery is? It’s voltage! In the same way, when our cells separate charge across the membrane, they create a voltage, or potential energy that can be used to do work. While all cells have a resting membrane potential, they are not the same for all cells. For example, the resting membrane potential for a neuron is -70mv while that of a red blood cell is only -10mv.This brings up an important concept: some cells are *excitable*, meaning they can be stimulated to experience a very predictable and rapid change in membrane potential. These are called action potentials and will be discussed later. Excitable cells include neurons, skeletal and cardiac muscle cells. Excitable cells have resting membrane potentials that range from -50mv to -85mv, while *non-excitable* cells have potentials ranging from -5mv to -10mv. Non-excitable cells do not experience action potentials and include blood cells and epithelial cells.

You might be thinking, “So what is an action potential?” Before we can fully dive into this concept, we need to revisit the concept of membrane transport proteins. Remember, membranes are made up of phospholipid bilayers that are virtually impermeable to any ionic or polar compound. Membrane transport proteins like channels (gated or leak) and carrier proteins embedded in the membrane enable the movement of these hydrophilic compounds across the membrane. You may remember that most channel proteins are very specific and will only allow certain things to pass through; some channel proteins are open all the time (**leak channels**) and others are closed most of the time (**gated channels**). Gated channels can be opened by several different types of stimuli. If they open in response to changes in membrane potential, they are called **voltage-gated ion channels**. Those that open in response to mechanical stimulation are called **mechanical-gated ion channels**, and those that open in response to a chemical signal are called **chemically-gated**, or **ligand-gated ion channels**. K^+ can move across the membrane through K^+ leak channels or gated channels. Na^+ can also move across the membrane via Na^+ leak and gated channels, but in most cells there are small

numbers of Na^+ leak channels (~1 Na^+ leak channel per 100 K^+ leak channels) and high amounts of Na^+ gated channels. Under normal resting conditions, Na^+ has a very large concentration gradient to enter the cell, but it cannot enter because the gated channel proteins for sodium are closed. In fact, for most ions, their specific protein channel is closed most of the time.

Activation of Voltage Gated Channels

If most gated ion channels are closed, how then are they stimulated to open in excitable cells? Proteins have many possible shapes or *conformations* that are dictated by the way in which their amino acids are arranged. Consider, for example, the following hypothetical voltage-gated ion channel.



Closed Voltage-Gated Channel. *Image created at JS, 2015.*

Depiction of a voltage-gated channel. This depiction shows how charges on amino acids can be attracted to or repelled by the "membrane potential" and contribute to the open or closed conformation of the "gates." The light blue "lever" represents the "activation gate" and the dark blue ball represents the "inactivation gate."

In the cartoon above, the voltage-gated ion channel actually has two gates depicted as a lever, which we will call the activation gate, and a ball, which we will call the inactivation gate. The lever, or activation gate, is connected to a helix or "spring" that "pulls" the lever towards the outer membrane causing the gate to open. Under normal resting conditions, this does not happen because the spring also has a net positive charge (notice the "+" symbols on the spring) which is repelled by the positive charges collecting on the outer membrane. At the same time, the positive charges of the spring are attracted to the negative charges collecting on the inner membrane. Thus, the activation gate remains in its closed conformation. If, however, we were able to change the charges, even slightly, on the membrane, then we could affect the conformation of the protein allowing the activation gate to open.

In the second membrane cartoon above, the charge was altered so that the inside suddenly became less negative and the outside less positive, notice that the spring could now recoil more towards the outer side of the membrane and "pull" open the activation gate (lever). Once open, ions specific to this channel could freely diffuse down their concentration gradients. However, the ball, or inactivation gate, also has charges on it that are attracted to complimentary charges in the mouth of the channel (not shown in the image). A few milliseconds after the activation gate (lever) opens, the inactivation gate (ball) closes the channel (see number 3 in the image above). The purpose of this inactivation gate is to regulate the amount of ions diffusing into the cell. If the channel were permeable to Na^+, then Na^+ could move in through the channel for a brief moment before the inactivation gate closed. Believe it or not, that

brief moment of Na^+ passing through the membrane is the basis of excitability. Does it appear clear now as to why these channels are called voltage-gated? It should make sense that changes in membrane voltage induce changes in protein channel shape which open and close the gate portions of the protein.

Movement of Ions Through Protein Channels

When an ion channel opens (even for <0.5 msec), the effect on the cell depends on the type of ion and the direction of diffusion. For example, since K^+ is high on the inside of the cell, the direction of movement, based on the concentration gradient, will be outward or towards the extracellular space. Since K^+ is a positively charged ion, moving more K^+ out will result in a loss of positive charges; hence, the cell will become even more negative. But wasn't K^+ already in equilibrium? Why would the movement of K^+ change if it was already in equilibrium by simply opening more channels? Well, the equilibrium of K^+ is based on the membrane permeability or the number of open channels. By opening voltage-gated K^+ channels, more K^+ will diffuse out of the cell than just the leak channels alone until a new electrical gradient is established to oppose the chemical gradient and a new equilibrium is established. Thus, the inside of the cell will become more negative because of the increased K^+ diffusion out of the cell.

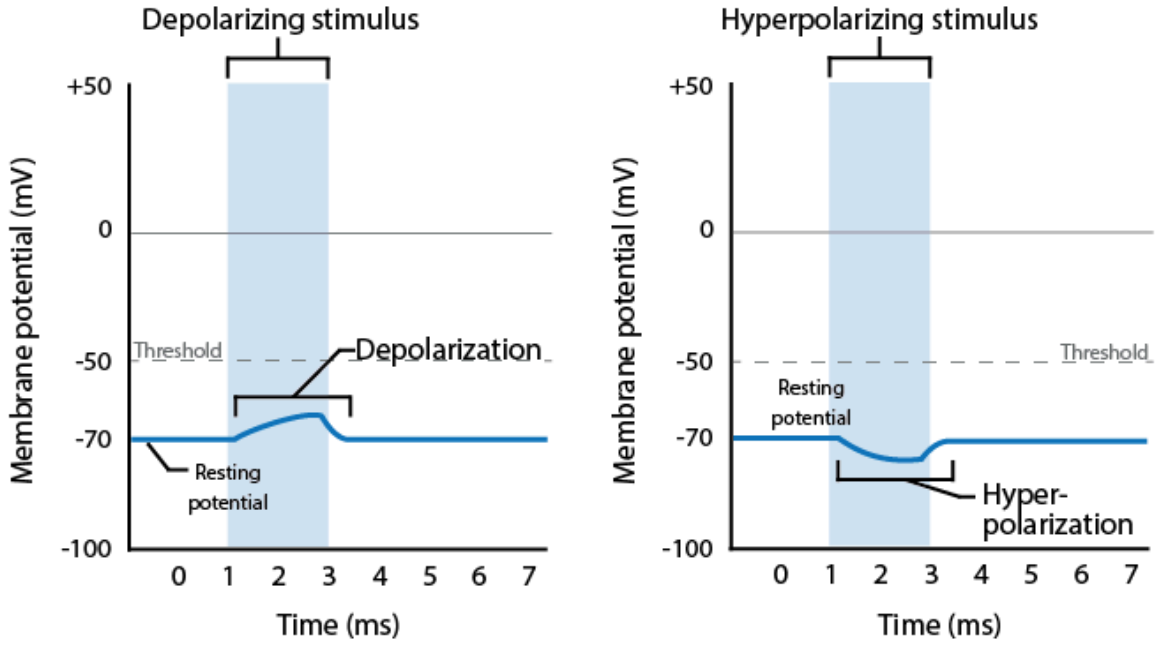
Another ion that can cause the membrane potential to become more negative is the Cl^- ion. In contrast to K^+, the concentration of Cl^- is highest on the outside of the cell, so the opening of a Cl^- channel will allow Cl^- to diffuse into the cell, also making the cell more negative. The opposite is true for Na^+ and Ca^{++}, both of which are positive ions and have concentrations that are highest on the outside of the cell. Thus, opening channels for either Na^+ or Ca^{++} will result in the inside of the cell becoming more positive as these cations diffuse into the cell. The following table shows the relative ion concentrations (mEq/l) in the intra- and extracellular fluids.

Ion	Extracellular fluid	Intracellular fluid
Na^+	142	10
K^+	4	140
Ca^{++}	2.4	0.0001
Mg^{++}	1.2	58
Cl^-	103	4
HCO_3^-	28	10

5.3.3

Graded Potentials

Let's cover some new terminology. Because we are dealing with charge differences and electrical currents (flow of ions), we use some unique terms to describe certain states of the membrane. At rest, the membrane is in a **polarized** state—polarized means there is a separation of charge across the membrane due to different ion concentrations on either side of the membrane (see table above). This polarized state is often referred to as the **resting membrane potential**. Already emphasized, the inside of the cell membrane will be negative in relation to the outside of the membrane. We can show this graphically by plotting membrane potential in mV on the y-axis and time in msec on the x-axis (see figure below). Any change in the membrane potential toward zero mV is termed a **depolarization** since the membrane is becoming less charged (i.e. there is less charge separation across the membrane). Any change in the membrane potential that moves back toward the resting potential is called a **repolarization**. And finally, any further decrease (more negative) in membrane potential below resting membrane potential is termed **hyperpolarization**. Note the prefixes of these terms as their meanings explain what is happening to the membrane potential. Opening channels for Na^+ or Ca^{++} would cause a depolarization, while opening channels for K^+ or Cl^- would cause a repolarization or even a hyperpolarization.



Graded Potentials. *Image created at BYU-Idaho, 2013.*
Graphical representation of *graded potentials*. The left graph shows a graded *depolarization*. Note that the membrane potential acutely increases (closer to 0mV) and then repolarizes to its resting membrane potential. The graph on the

right shows a graded *hyperpolarization* as the membrane potential acutely becomes more negative than resting membrane potential.

Graded potentials are small in magnitude, meaning they don't drastically change the membrane potential. The magnitude of the graded potential is proportional to the strength of the stimulus. Hence, a strong stimulus might result in a 10mV change in the membrane potentials, while a weaker stimulus may produce only a 5mV change. Graded potentials are also localized to a small area of the cell membrane and are sometimes called **local potentials**. Graded potentials are fast-acting, meaning that the membrane potential typically returns to resting membrane potential quickly, within a couple msec. The opening of mechanical or ligand-gated ion channels are what induce graded potentials. Something unique to graded potentials is that they can be summed (added together) to increase the change in membrane potential. For example, if a depolarizing stimulus is repeated over and over in a short period of time, it can result in an even larger depolarization as each subsequent graded potential further depolarizes the membrane before it can fully repolarize back to the resting membrane potential. If a strong depolarizing graded potential (or multiple depolarizing graded potentials) increase the membrane potential above the **threshold potential** for the cell, an action potential will occur. This is the great significance of graded potentials: They determine whether action potentials occur or not. You can think of graded potentials as the triggers that initiate action potentials; without them, action potentials would not happen. Let's now discuss action potentials.

5.3.4

Action Potentials

An action potential is where things really become interesting and exciting—no pun intended; but remember that only excitable tissues can experience action potentials. An **action potential** is simply a rapid and drastic depolarization of the membrane potential followed by a rapid repolarization to the resting membrane potential (see figure below). Unlike graded potentials, action potentials are not localized, but propagated throughout the entire cell membrane. The action potential is the basis for transmitting signals in nerve cells, inducing muscle contraction, and perception of all our senses. Action potentials are initiated when depolarizing graded potentials reach **threshold potential**. This is the specific membrane potential that induces activation of the **voltage-gated ion channels** responsible for action potentials and is most often the voltage-gated Na^+ channel. If a graded potential is not strong enough to bring the membrane potential up to threshold, it is called a sub-threshold stimulus. On the other hand, if threshold potential is met, or even exceeded, an action potential will result. This phenomenon is known as the "all or nothing principle." Once an action potential is initiated, it will be the same predictable depolarization followed by repolarization; there are not different sizes of action potentials; hence, **ALL** of an action potential or **NOTHING**. Even if threshold potential is greatly exceeded by a super-large graded potential, an action potential of equal magnitude to any other action potential experienced by that cell is initiated. Unlike graded potentials, action potentials cannot be summed or added upon.

In a neuron at rest, there is very little diffusion of Na^+ across the membrane (very few Na^+ leak channels). However, if the cell membrane of the neuron experiences a graded potential (or multiple graded potentials) that is sufficient to depolarize the membrane to threshold potential (approximately -55mV), an action potent will initiate as the voltage-gated Na^+ channels change conformation allowing the "activation gate" to open (Review the section above on activating voltage-gated ion channels). Because the concentration of Na^+ is extremely high on the outside of the cell, the opening of Na^+ channels will cause a rapid influx of Na^+ down its concentration gradient, further depolarizing the membrane as more positive Na^+ ions attempt to diffuse into the cell. The membrane potential will increase to almost +30mV when the inactivation gate closes and no further Na^+ diffusion occurs. It is important to note that depolarization occurs with minimal changes in the overall concentration of Na^+ or K^+ (Only one out of every 100,000 Na^+ ions need to enter the cell to produce a 100mV change in membrane potential).

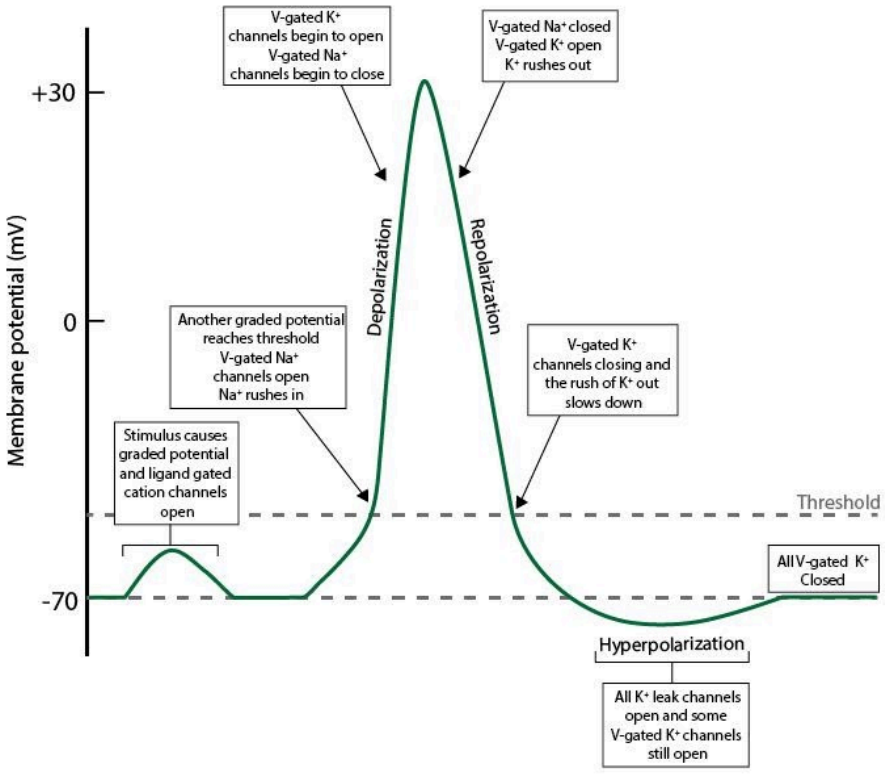
The activation and inactivation gates of the voltage-gated Na^+ channels are unique to this channel's functions: The activation gate is very sensitive to voltage changes and is the basis of threshold (as described above). The inactivation gate is less sensitive to voltage, and thus slightly delayed compared to the activation gate, which allows the channel to transport Na^+ for a brief moment. As the membrane potential increases due to the activated voltage-gated Na^+ channels, voltage-gated K^+ channels begin to open resulting in an increased efflux of potassium out of the cell. This efflux, in addition to the efflux resulting from K^+ leak channels, is responsible for repolarization and even hyperpolarizing the membrane. There are many different types of voltage-gated K^+ channels expressed in neurons, some of which are activated at different membrane voltages. However, unlike voltage-gated Na^+ channels, voltage-gated K^+ channels only have activation gates. Because these voltage-gated K^+ channels lack inactivation gates, they are slower to close during repolarization. As a result, every neuronal action potential features a **zone of hyperpolarization** due to the additional diffusion of K^+ through voltage-gated K^+ channels that have yet to close AND leak channels. Once the voltage-gated K^+ channels close, the membrane will return to the resting potential established by the K^+ leak channels. The small diffusions of K^+ and Na^+ during each action potential are then reestablished by the Na^+/K^+ ATPase pump, however this

is not necessary for another action potential. In fact, it has been demonstrated that the ion gradients within a neuron are sufficient to generate 10,000 action potentials without replenishment from the Na^+/K^+ ATPase pump.

Here are two videos about the resting membrane and action potentials to aid you in your learning and understanding:

https://books.byui.edu/-eEkp

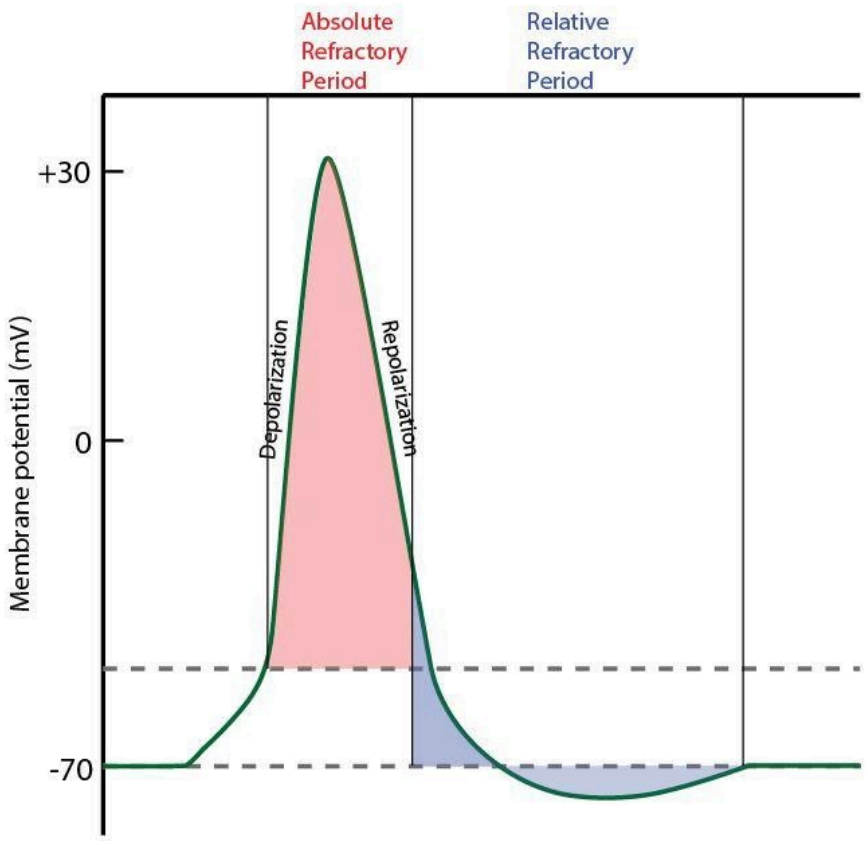
https://books.byui.edu/-Bnd



Action Potential. *Image created by BYU-Idaho student, Kaylynn Loyd 2013*

5.3.5

Refractory Periods



Refractory Period. *Image created at BYU-Idaho, Fall 2015*

Another concept to be discussed is the **refractory period**. By definition, the refractory period is the amount of time during which a cell is incapable of repeating another action potential after one has been initiated. There are two types of refractory periods: The **absolute refractory period**, which is the interval of time during which a second action potential cannot be initiated, no matter how large a stimulus is repeatedly applied. Second, the **relative refractory period**, which is the interval of time during which a second action potential can be initiated, but initiation will require a greater stimulus than before. Refractory periods are caused by the inactivation gate of the voltage-gated Na^+ channel. Once inactivated, the Na^+ channel cannot respond to another stimulus until the activation and inactivation gates are reset.

Here is a video to help with understanding:

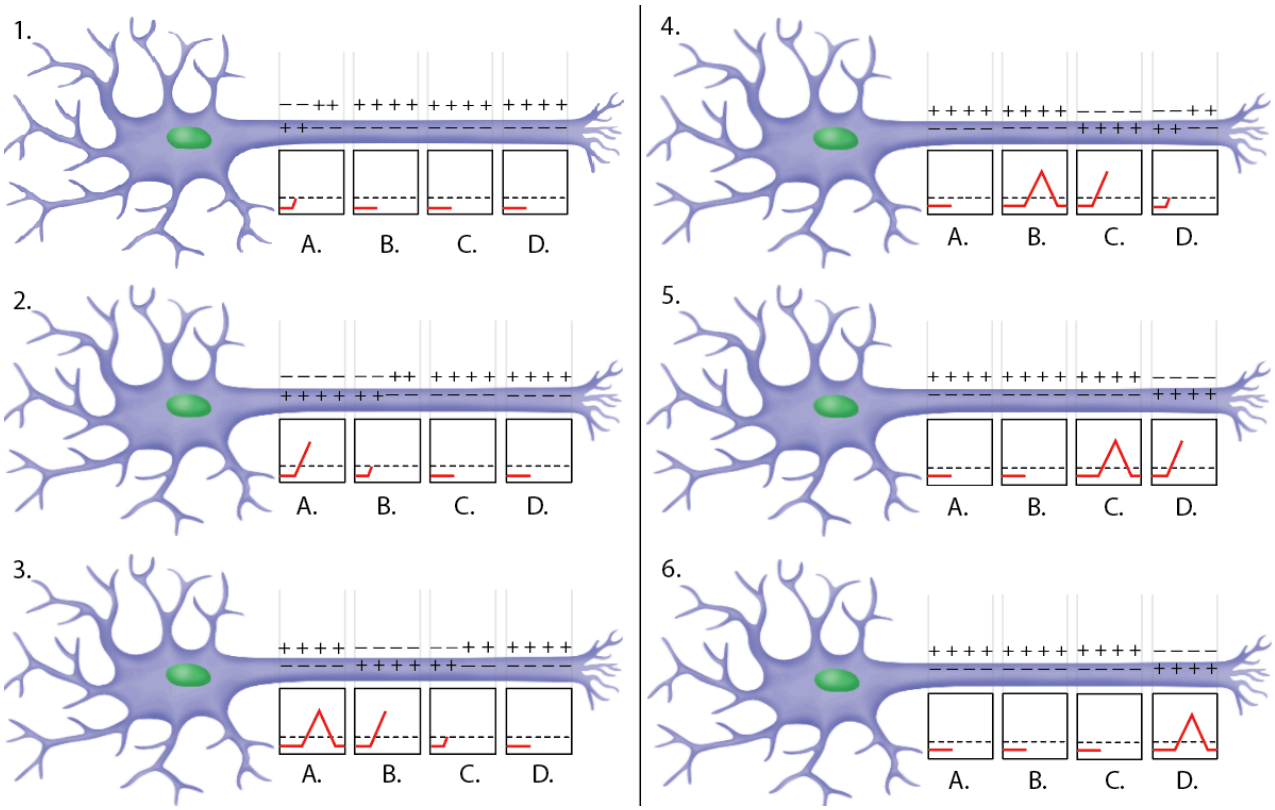
https://books.byui.edu/-KoSc

5.3.6

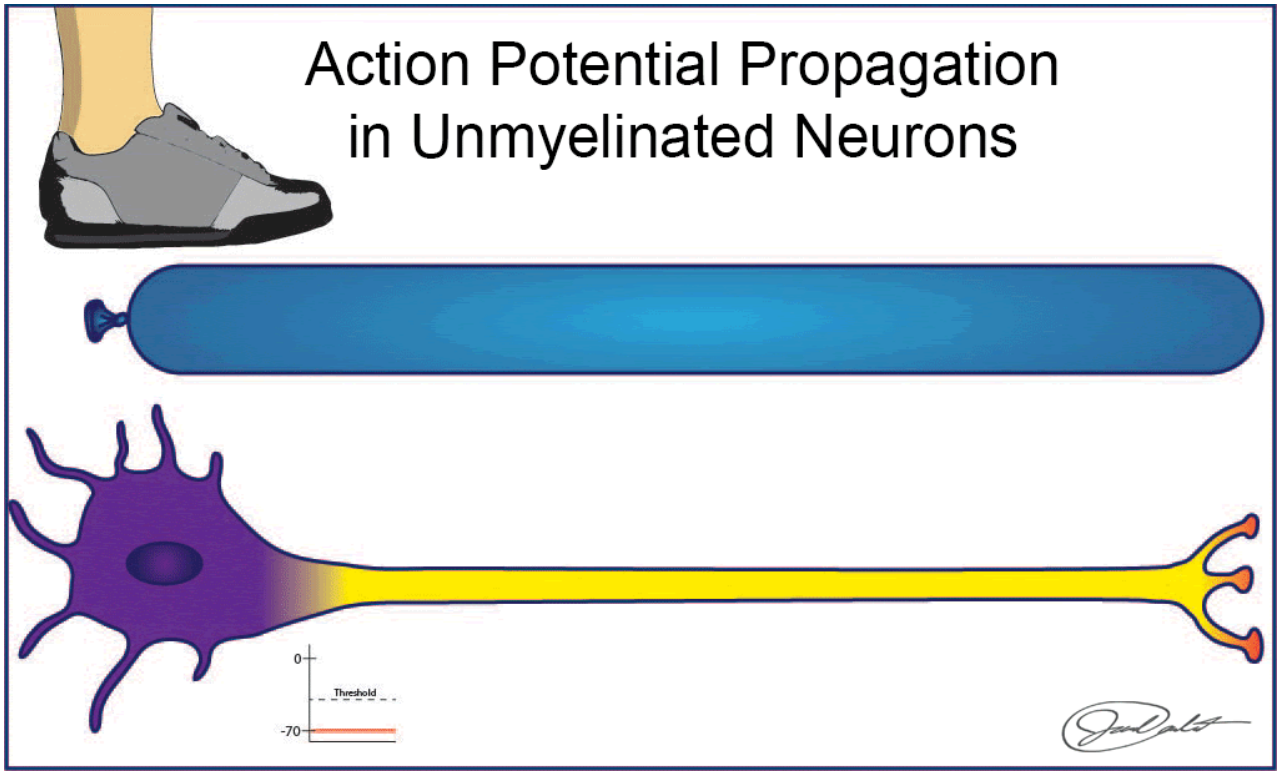
Propagation of an Action Potential

Action potentials are usually generated at one end of a neuron, typically the cell body, or soma, and then "propagated" like a wave along the axon towards the opposite end of the neuron.

The image below shows how an action potential might have started near the cell soma (notice the depolarization in 1A) and as it propagates down the axon towards the opposite end (2-6). The membrane potential behind the moving action potential eventually repolarizes and returns to resting membrane potential (e.g. 4A). The axon ahead of the depolarization current has not yet depolarized, and it is also at resting membrane potential (e.g. 2C&D). Where the action potential is occurring, we find the membrane potential depolarized, and the outside of the membrane at that spot is negatively charged relative to the inside of the membrane at that spot (e.g. 3B). As sodium diffuses into the cell, it will depolarize the next adjacent spot on the axon in the direction that the action potential is propagating (e.g. 3C). Action potentials cannot propagate in reverse because that section of membrane is most likely in refractory periods and does not depolarize.

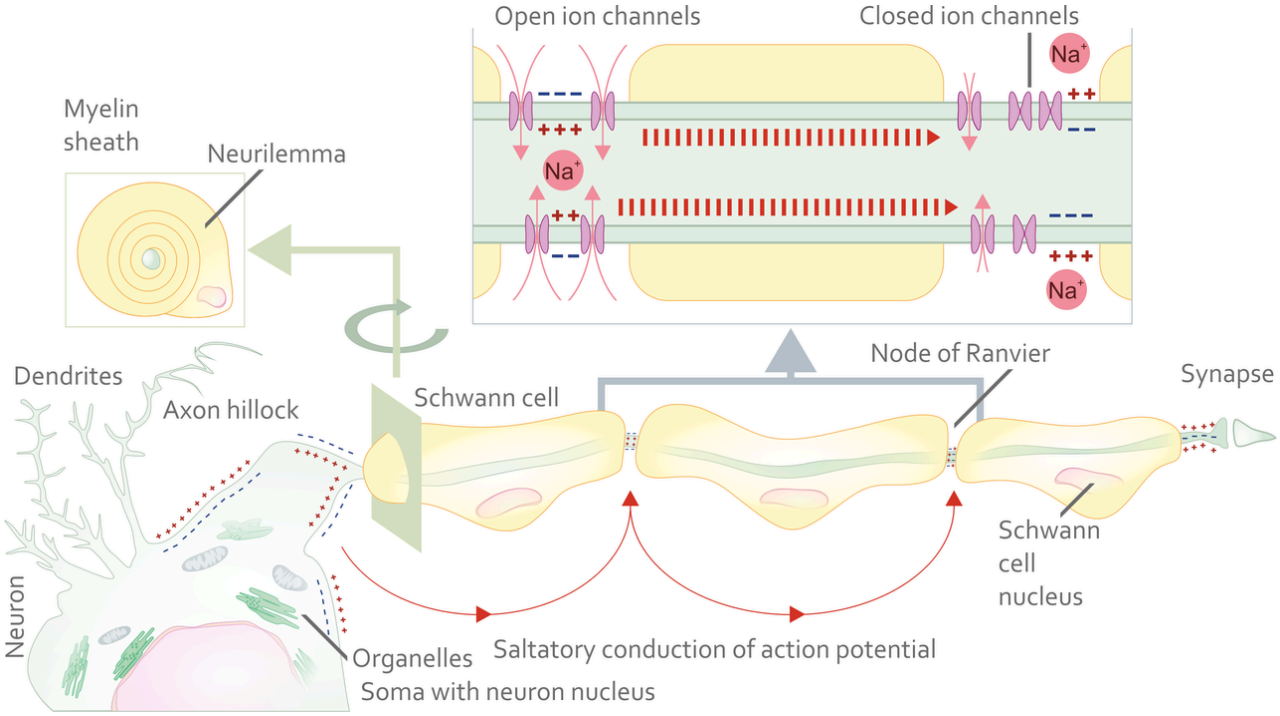


The image above shows how an action potential propagates down the neuron axon. Each number (1-6) represents a different timepoint of action potential propagation. The square insets below each neuron shows the stage of an action potential that exists at each axon segment. *Image by Beck Torgerson S18*



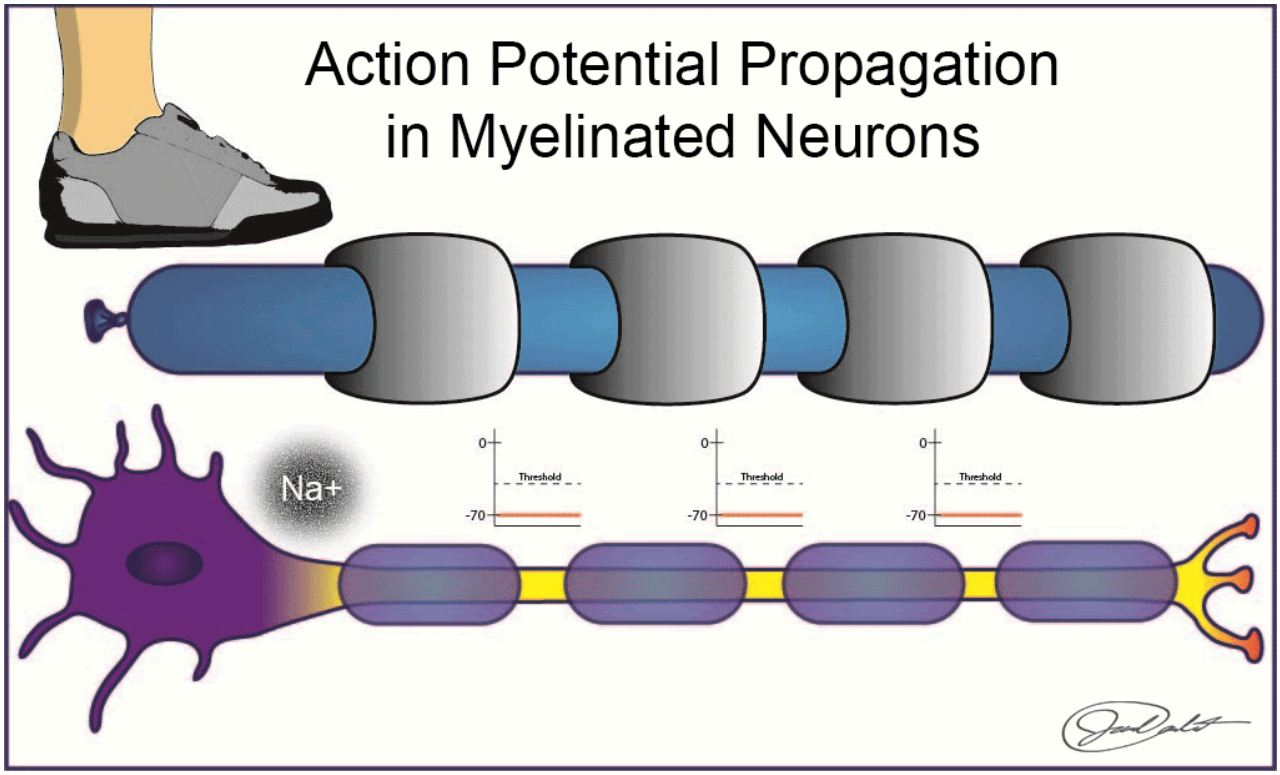
Animation of Action Potential Propagation in Unmyelinated Neurons. *BYU-Idaho image created Winter 2015*

The image above is another animation that you can watch if you click on the picture. This animation shows how an action potential traveling down the axon is similar to stepping on one end of a water balloon. In reality, a pressure wave in the water balloon would get smaller as it traveled down the length, but a traveling action potential (or depolarization wave) is recreated at every spot on the axon that has voltage-gated sodium channels that open at threshold potential. In this way, the original strength of the depolarization wave is continually recreated.



Propagation of Action Potential Along Myelinated Nerve Fiber. *Author: Helixitta; Site: https://commons.wikimedia.org; /wiki/File:Propagation_of_action_potential_along_myelinated_nerve_fiber_en.png; License: This file is licensed under the Creative Commons Attribution-Share Alike 4.0 International license.*

The image above shows a myelinated peripheral nerve axon. Myelin is a lipid and protein rich tissue that wraps around the axon in a way that "insulates" it from depolarization. In the peripheral nervous system, myelin is made by Schwann cells. Non-myelinated sections in between Schwann cells are called *nodes of Ranvier*. These sections can experience depolarizations and action potentials. What is the effect of myelin? It serves to increase action potential conduction (propagation) time. Since action potentials can only occur at the nodes of Ranvier, action potential propagation can travel much faster (nearly 10 times faster than unmyelinated axons) because depolarizations 'jump' from node to node. This phenomenon is known as **saltatory conduction**. *Saltatory* means, 'jumps or leaps."



Animation of Action Potential Propagation in Myelinated Neurons. *BYU-Idaho image created Winter 2015*

The image above is another animation (click on the picture). It shows how a myelinated axon might compare to a water balloon with segmented cuffs on it. Upon stepping onto the water balloon, a pressure wave is generated that is recreated at each "node;" although notice that, like before, the magnitude of the pressure wave decreases along the length of the water balloon. Similarly, when a myelinated neuron initiates an action potential, the positively charged sodium entering in at the axon hillock (area closest to the soma) causes positive charges to rapidly travel down the axon where they can depolarize each node. Like the water balloon, the strength of these depolarizations decrease along the length of the axon. However, at the most proximal node of Ranvier, this depolarization is sufficient to reach threshold and an action potential is re-created. This re-created action potential then propagates stronger depolarizing currents to the next most proximal node which re-creates another action potential at that node. This process continues along the length of the axon and enables faster action potential conduction time. Consider these three things:

1. The original depolarization event at the axon hillock will open sodium channels at nodes of Ranvier that can facilitate those membrane sections getting closer to threshold.
2. Each node that reaches threshold re-creates an action potential (depolarization wave) that is equal to the first.
3. Depolarization occurs only on bare axon between myelin segments and not along the entire axon surface. This is saltatory conduction and is how myelination increases action potential conduction speed.

SUMMARY

Let's see how well we understand the concepts of this chapter. Consider your fingertips; there are at least five different types of touch receptors that allow you to feel various textures and pressures, but how do they work? Touch receptors are just fancy neurons, but they exhibit the same kinds of phenomena that we just talked about. At rest, these sensory neurons are permeable to K^+ but not Na^+, so the inside of the membrane is negative relative to the outside (-70mV). Consequently, voltage-gated Na^+ channels are in a closed conformation. In order for us to sense touch, we need to convert the touch stimulus into an action potential, but how? The mechanical stimulus of touch causes a conformation change in mechanically-gated Na^+ channels, causing them to open. As Na^+ diffuses into the cell, the positive charges depolarize the membrane (graded potentials) to threshold potential which opens voltage-gated Na^+ channels and initiates an action potential. This action potential is propagated to the brain where it is perceived as touch. Believe it or not, every external stimulus—whether taste molecules, light waves, sound waves, or mechanical touch—is converted to an action potential. Action potentials are a main way the body communicates between tissues and the brain works strictly by discerning and initiating action potentials. These principles will be applied and detailed in later chapters, so please make sure you understand the basics of action potentials, electrophysiology, and membrane transport.

6.0

MODULE 6: NERVOUS SYSTEM ORGANIZATION

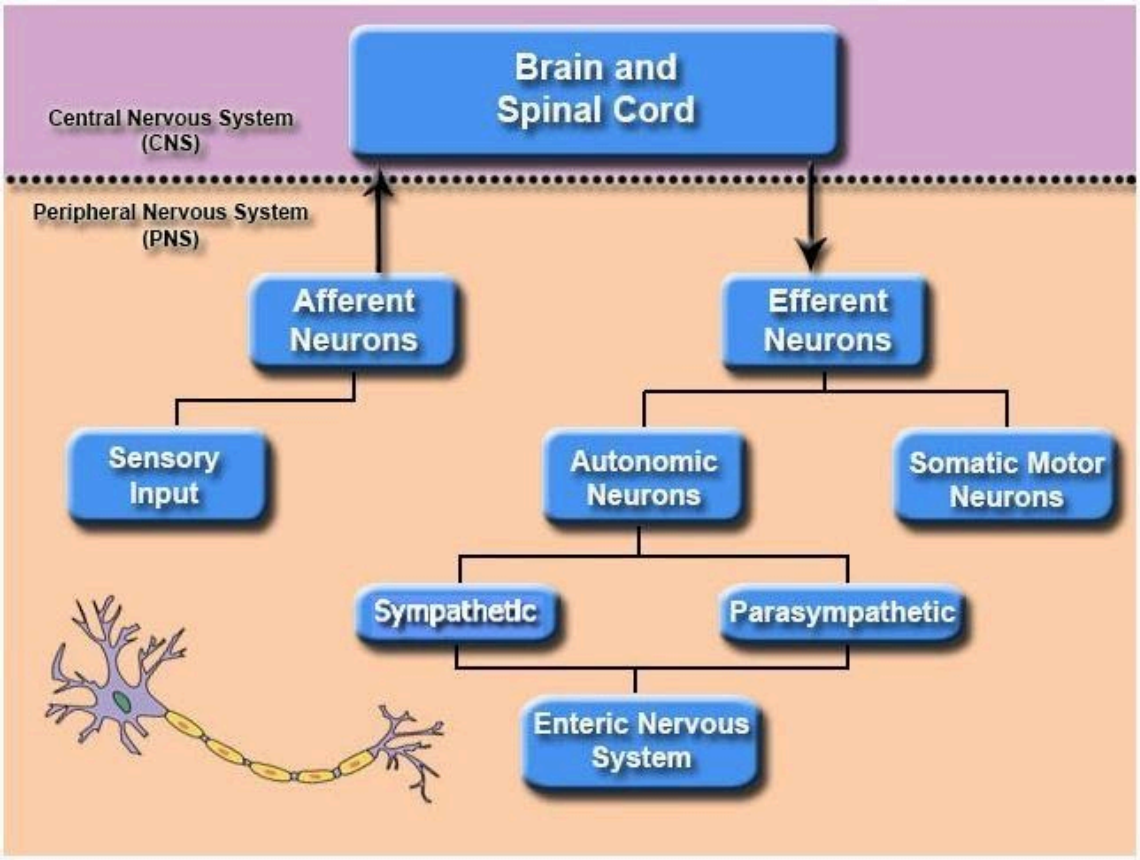
Imagine that you suddenly lost the ability to stand unless you looked down at your feet. Image if your arms seemed to wander unless you kept an eye on them. This is the exact situation that a woman named Christina found herself in after waking up one morning. She exclaimed, “Something awful has happened, I can’t feel my body. I feel weird—disembodied.” An integral part of her nervous system had suddenly deteriorated due to what was later diagnosed as polyneuritis, an inflammatory disorder that affects the peripheral nervous system and causes loss of myelin. Essentially, overnight, Christina had lost all **proprioception**: the ability to sense the relative position of body parts. Sense of body positioning is determined by three things: vision, balance organs, and proprioception. Christina lost proprioception, and because of it, she had to learn to control her body with her eyes. She found that she could do nothing without using her eyes. In fact, her body would collapse into a heap the minute she closed her eyes. Gradually over time, Christina learned to walk again and to function with the usual business of life but only with great care in maintaining attention to the particular movement and never at the same level as before. She found that there was no in-between or gradual change with movement. Strict focus had to be maintained for even the simplest tasks. In her words, she stated, “I feel my body is blind and deaf to itself...it has no sense of itself” (Sacks, 1985). In this section, you will learn about the nervous system, how it controls the body, and perhaps more importantly, as illustrated in Christina’s case, how the nervous system interprets the environment around us.

6.1

ORGANIZATION OF THE NERVOUS SYSTEM

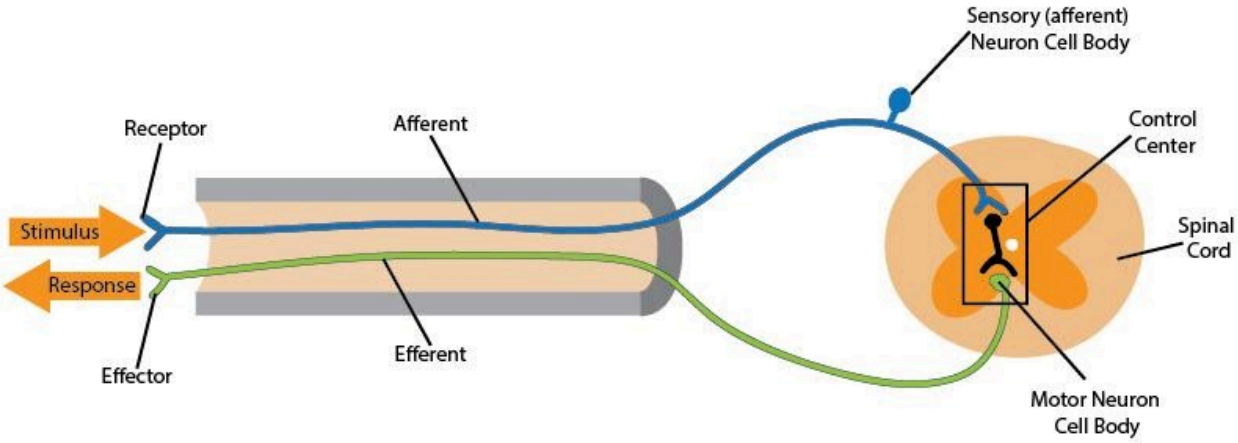
The nervous system coordinates voluntary and involuntary actions in the body by sending and receiving information. The nervous system is comprised of an enormous number of cells (over 100 billion), primarily of two types: **neurons** (the signaling units) and **glial cells** (the supporting units). However, nervous system function is mostly a story of the neuron. The neuron is the functional unit of the nervous system and is designed to transmit information between cells. Interestingly, neurons with a particular function are found in a predictable location. This regularity in structure has permitted neurobiologists to categorically organize the nervous system based on location and function (see figure below).

Thus, the nervous system can first be divided into two major parts: the central nervous system (CNS) and the peripheral nervous system (PNS). The CNS consists of neurons associated with central processing and which are located in the brain and spinal cord. The *peripheral nervous system* (PNS) consists of neurons associated with sensory input (afferent) and motor output (efferent) and functions to connect the central nervous system to all other parts of the body. Stated another way, if the entire structure of the neuron is contained within the brain and/or spinal cord, the neuron would be considered part of the CNS. In contrast, if any part of the neuronal structure is located outside of the brain and/or spinal cord, the neuron would be considered part of the PNS (see image below).



Organization of the Nervous System. *Image generated at BYU-Idaho Spring 2013*

For the most part, information is transmitted between these two systems following this basic pattern: stimulus, receptor, afferent pathway (input signal), control center, efferent pathway (output signal), effector, and response. In other words, sensory receptors located throughout the body constantly monitor the conditions of the environment and send this information via the PNS to the CNS for central processing. If a response is needed (i.e. to maintain homeostasis), the CNS will send new information through the PNS to target organs that will help adjust to the initial stimulus. It should be noted that some functions can be contained entirely within the CNS: for example, dreaming, thinking, or even information storage.



Peripheral Nervous System (PNS) Communication with Central Nervous System (CNS).
Image generated at BYU-Idaho Spring 2013

Neurons of the efferent division of the PNS can be further subdivided into the somatic nervous system, which controls the voluntary movement of skeletal muscle, and the autonomic nervous system, which regulates involuntary functions of organs (such as the heart, lungs, glands, etc.) and smooth muscle tissues (airways, blood vessels, etc). Autonomic neurons are further subdivided into sympathetic and parasympathetic systems (see first figure). The autonomic nervous system will be addressed in a separate module.

A third division of the PNS is a semi-independent nervous system called the *enteric nervous system*, which controls the gastrointestinal tract. This system is considered semi-independent because it can run independently or through modulation by the autonomic nervous system, particularly from the parasympathetic system (see first figure). It is also interesting to note that the enteric nervous system contains more neurons than the entire spinal cord.

6.1.1

Neuron Structure and Classification

Neurons have four specialized structures that allow for the sending and receiving of information: the cell body (soma), dendrites, axon, and axon terminals (see figure below, lower left).

Cell body or Soma

The cell body is the portion of the cell that surrounds the nucleus and plays a major role in synthesizing proteins.

Dendrites

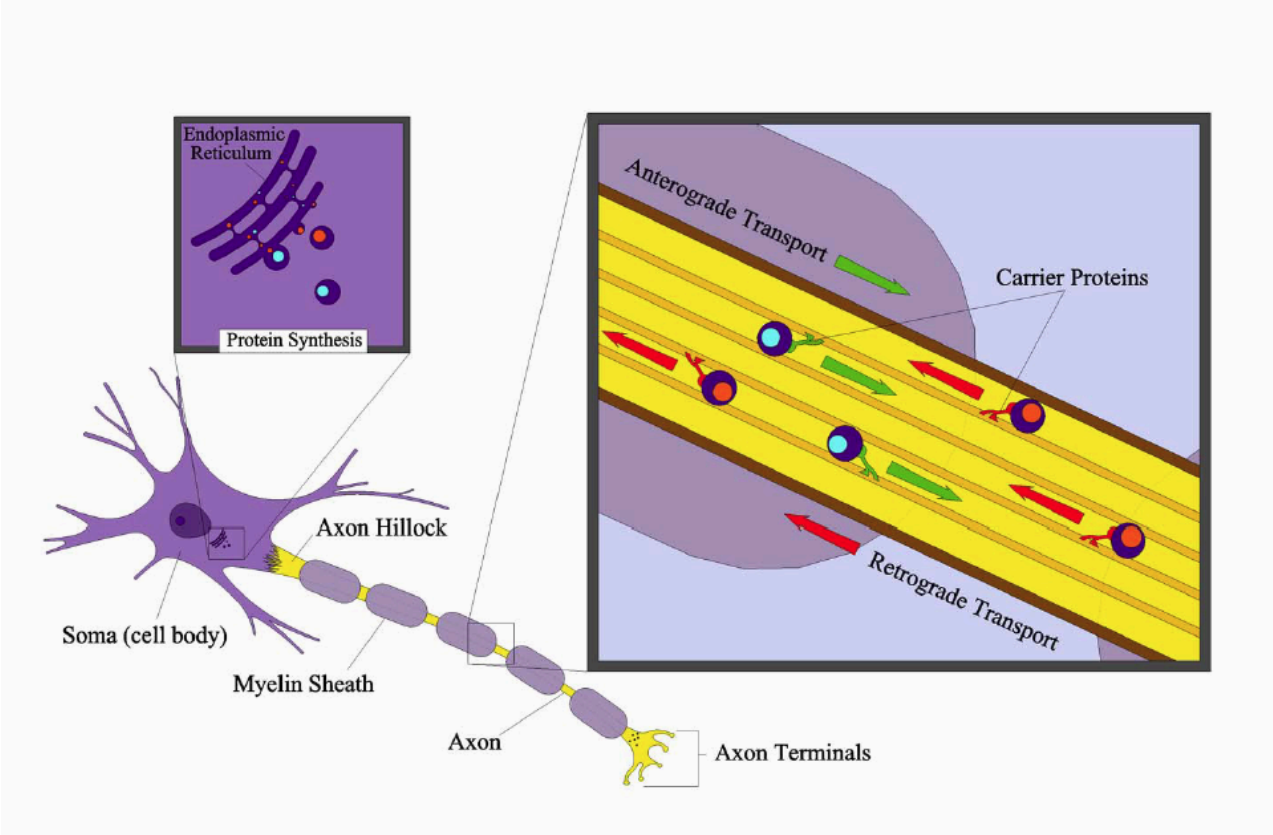
Dendrites are short, branched processes that extend from the cell body. Dendrites function to receive information and do so through numerous receptors located in their membranes that bind to chemicals called neurotransmitters. Many of these receptors are lignad-gated channels.

Axon

An axon is a large process that extends from the cell body at a point of origin—called the axon hillock—and functions to send information. In contrast to the shorter dendrites, the axon can extend for more than a meter. Because of this length, the axon contains microtubules and is surrounded by myelin. Microtubules are arranged inside the axon as parallel arrays of long strands that act as highways for the movement of materials to and from the soma. Specialized motor proteins "walk" along the microtubules, carrying material away from the soma (**anterograde transport**) or back to the soma (**retrograde transport**). This system can move materials down the axon at rates of 400 mm a day (see lowest figure). Myelin consists of totally separate cells that coil and wrap their membranes around the outside of the axon. These are essential for electrical insulation and to speed up action potential propagation.

Axon Hillock

As described above, the axon hillock is the origin of an axon. It has many voltage=gated Na+ channels. When thes channels reach threshold an action potential is initiated which will the propagate down the zxon to the axon terminal. In essence, the is region is when a neuron “decides” whether or not to have an ation potential.



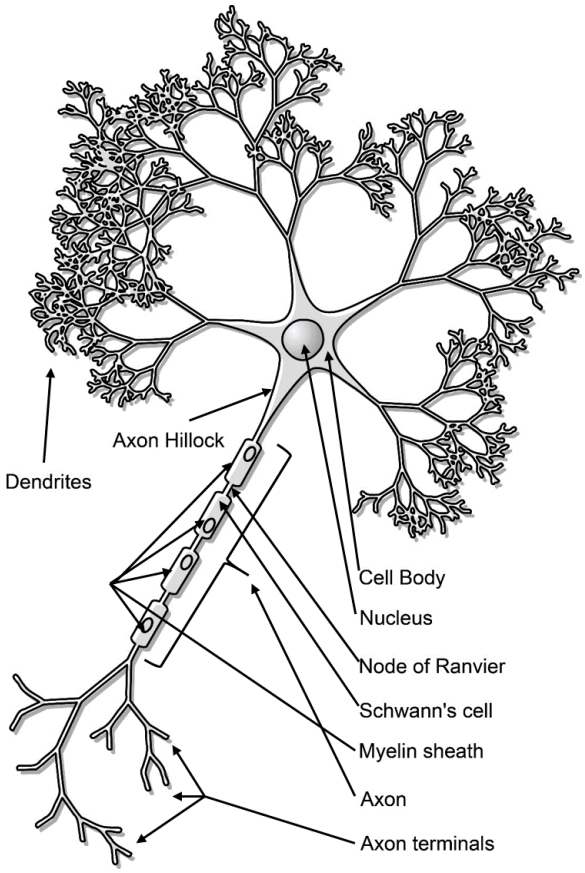
Anterograde and Retrograde Transport in an Axon. *Image produced by BYU-Idaho Student Jared Cardinet 2013*

Axon terminals

Where an axon reaches a target, it terminates into multiple endings, called axon terminals. The axon terminal is designed to convert the electrical signal of an action potential into a chemical signal to be sent to a neighboring cell in a process called synaptic transmission (further explained in the section "Physiology of the Neuron"). This region contains many voltage-gated Ca2+ channels that play a critical role in synaptic transmission.

Neuron Life Cycles

Most neurons are amitotic (lose their ability to divide). Exceptions to this rule are found in olfactory neurons (those associated with smell) and hippocampal regions of the brain (those associated with memory). Fortunately, lifespans of amitotic neurons is near 100 years. Still, if a neuron is damaged or lost, it is not easily replaced. For this reason, there is usually limited recovery from serious brain or spinal cord injuries. Perhaps the slow recovery rate or lack of regeneration is to ensure that learned behavior and memories are preserved throughout life. Neurons also have exceptionally high metabolic rates and subsequently require high levels of glucose and oxygen. The body will go to great lengths to ensure that neurons are adequately fed; in fact, if for some reason the brain detects that it is not receiving adequate amounts of nutrition, the body will shut down immediately (i.e. faint).



Key Neural Structures.
Title: Neuron-figure-notext.svg; Author: Nicolas.Rougier;

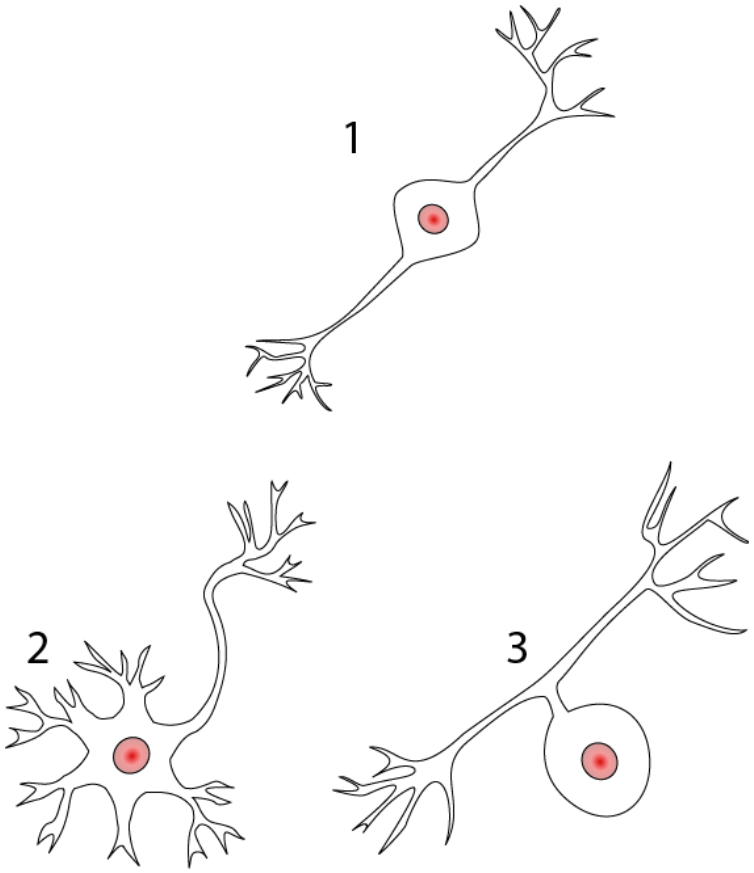
Site: https://books.byui.edu/-jtcV

License: This file is licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license.

Structural Classification of Neurons

Neurons can be classified base up their structure or morphology. Structural classification of neurons is based upon the number of processes that extend out from the cell body. Three major groups arise from this classification: **multipolar, bipolar,** and **unipolar** neurons.

Multipolar neurons are defined as having three or more processes that extend out from the cell body. They comprise more than 99% of the neurons in humans and are the major neuron type found in the CNS and the efferent division of the PNS.



Structural classification of neurons. 1) Bipolar 2) Multipolar 3) Unipolar.
Title: Neurons uni bi multi pseudouni.svg; Author: Pseudounipolar_bipolar_neurons.svg: Juoj8 derivative work: Jonathan Haas;

Site: https://books.byui.edu/-VIRq;

License: This file is licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license.

Bipolar neurons have only two processes that extend in opposite directions from the cell body. One process is called a *dendrite*, and another process is called the *axon*. Although rare, these are found in the retina of the eye and in the olfactory system.

Unipolar neurons have a single, short process that extends from the cell body and then branches into two more processes that extend in opposite directions. The process that extends peripherally is known as the *peripheral process* and is associated with sensory reception. The process that extends toward the CNS is the central process. Unipolar neurons are found primarily in the afferent division of the PNS.

Functional Classification of Neurons

Neurons can also be classified functionally according to the direction in which they transmit signals, in relation to the CNS. This classification also results in three different types of neurons: **sensory neurons, motor neurons**, and **interneurons**.

Sensory neurons, or afferent neurons, transmit information from sensory receptors in the skin or the internal organs toward the CNS for processing. Almost all sensory neurons are unipolar.

Motor, or efferent, neurons transmit information away from the CNS toward some type of effector. Motor neurons are typically multipolar.

Interneurons are located between motor and sensory pathways and are highly involved in signal integration. The vast majority of interneurons are confined within the CNS.

6.1.2

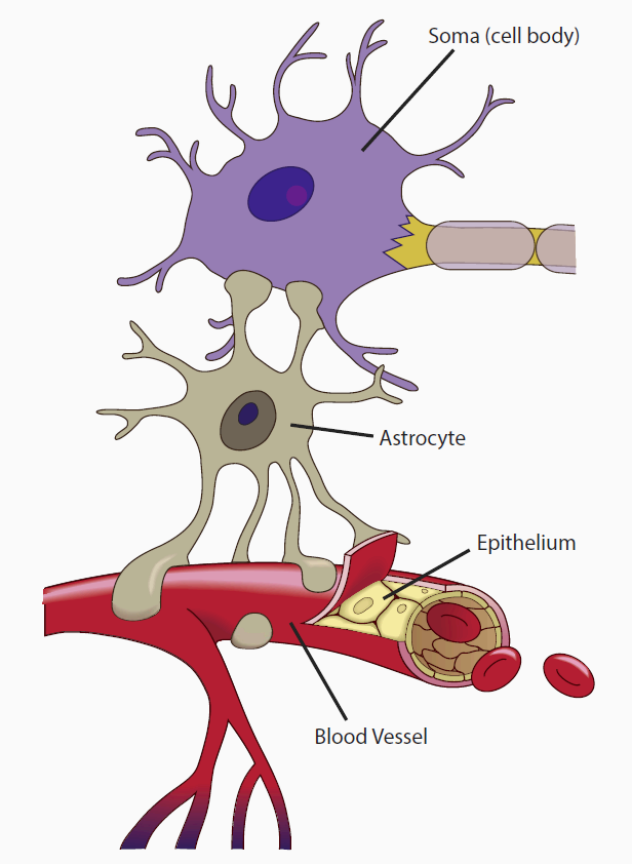
Glial Cells of the CNS

Unlike neurons, the glial cells can be replaced if they are damaged. Glial cells compose half of the volume of the brain and are more numerous than neurons. There are four major types of glial cells in the CNS: the astrocyte, the oligodendrocyte, the ependymal cell, and the microglial cell.

Astrocyte

Astrocytes have an enormous amount of processes that wrap around blood vessels and neurons. Because of this arrangement, astrocytes are ideally positioned to control and modify the extracellular environment around neurons. Most of the functions of the astrocytes are attributed to controlling this environment. One of the most important ways that astrocytes control this environment is by contributing to the blood-brain barrier formed by blood vessel endothelial cells in the CNS. The blood-brain barrier is a highly selective membrane that tightly controls what solutes, chemicals, pathgoens, etc. can enter the CNS.

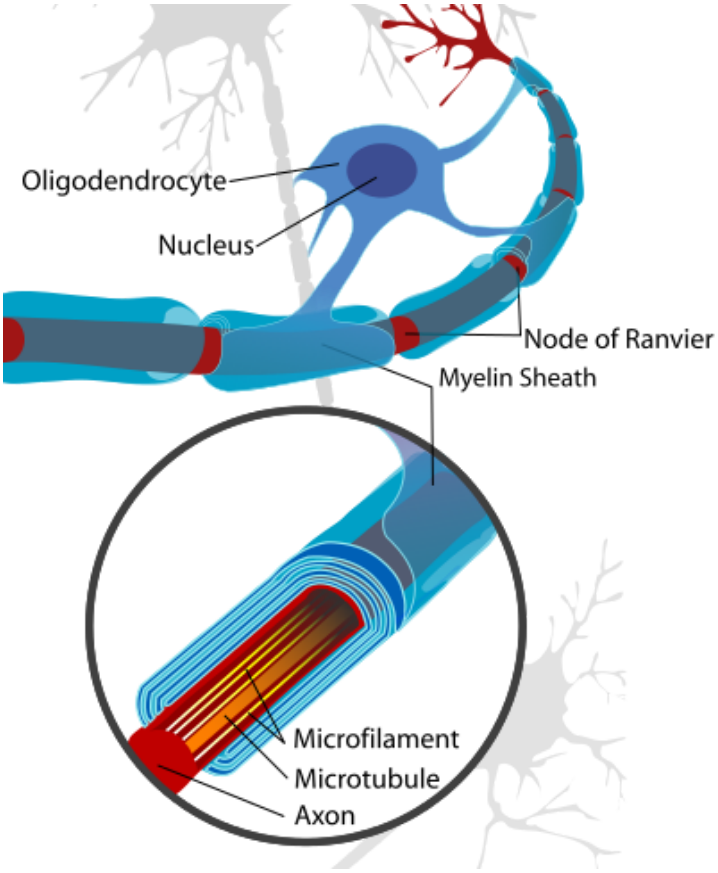
Astrocyte characteristic	Function
Glycogen storage	Astrocytes store all the glycogen present in the CNS. This glycogen is used to help meet the high metabolic needs of the CNS. The main source for these metabolic needs is blood glucose, but glycogen levels can sustain the needs of the CNS for 5–10 minutes.
K^+ permeability	Active neurons lose K^+ into the extracellular spaces, which would act as a positive feedback system for depolarization if the K^+ was not trapped by the astrocytes. They take up K^+ by a pump (Na^+/K^+ ATPase pump) and co-transporters ($Na^+/K^+/Cl^-$ and K^+/Cl^- exchangers).
Gap Junctions	Astrocytes are coupled to each other, as well as other glial cells and neurons, through gap junctions. This may serve to help modulate activity and sensitivity in the CNS.
Neurotransmitters	Astrocytes synthesize over 20 different neurotransmitters and take up excess neurotransmitters to help terminate signals at the synapse.
Growth factors	Astrocytes secrete a variety of growth factors, which are important for the establishment of fully functioning excitatory synapses.
Blood flow	Astrocytes can modulate blood flow in the brain by inducing localized vasodilation or vasoconstriction. This modulation can occur through gap junctions between the astrocytes and the endothelial cells of brain blood vessels.



Astrocyte Processes Associated with Capillaries and Neurons.
Image by BYU-Idaho student, Jared Cardinet, 2013.

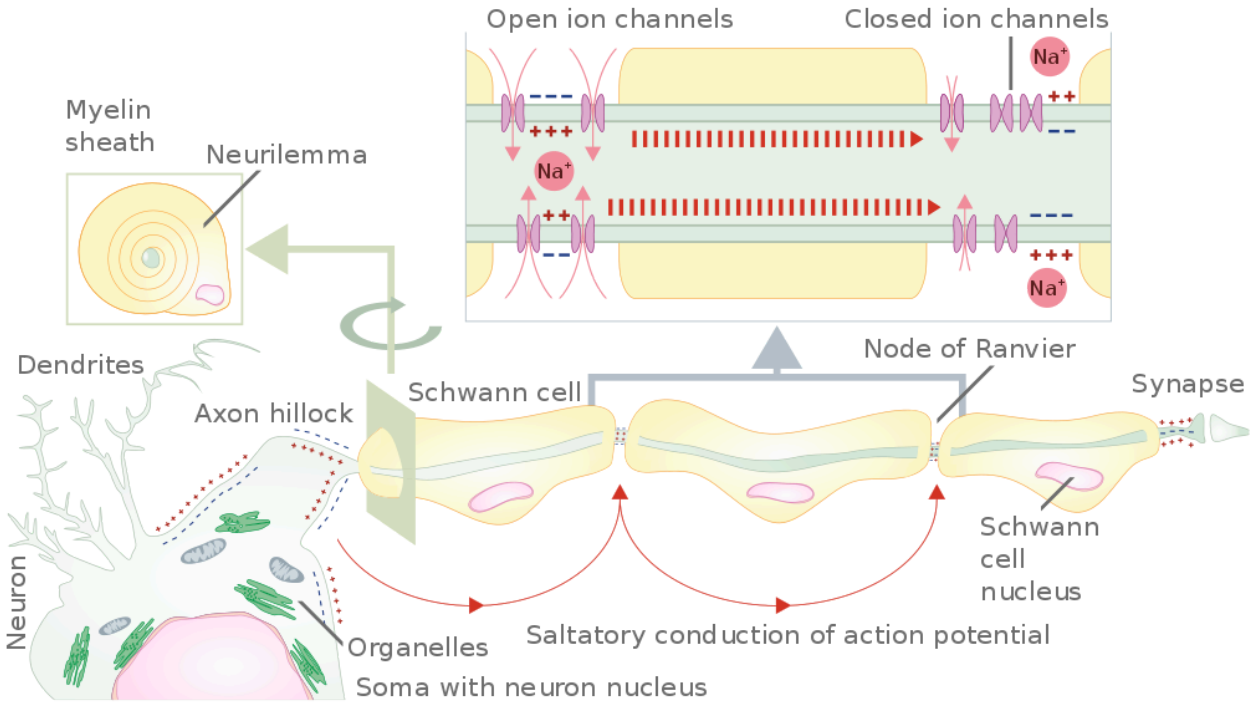
Oligodendrocyte

The primary function of the oligodendrocyte is to provide and maintain the myelin sheaths around axons in the CNS. Myelin is the insulating component of the nervous system. It allows for electrical signals to be propagated along one axon without being spread to other axons. Oligodendrocytes send out 15–30 long processes, which wrap many times around a section of an axon. Between each "wrapping," there is a small area of exposed axon called the **node of Ranvier.**



Oligodendorcyte. *File: Neuron_with_oligodendrocyte_and_myelin_sheath.svg: *Complete_neuron_cell_diagram_en.svg: Author: LadyofHats derivative work: Andrew c; Liscence: [Public domain], via Wikimedia Commons*

The wrapping creates many layers of tightly compressed membranes that is called **myelin**. Myelination speeds up the conduction of action potentials down the axon by allowing the action potentials to occur only at the nodes, a process called *saltatory conduction*. Myelination also induces the clustering of voltage-gated Na^+ channels at the nodes. In addition to myelination, oligodendrocytes also play key roles in pH regulation of the CNS.



Propagation of Action Potential Along Myelinated Nerve Fiber.*Author: Helixitta; Licence: [CC BY-SA 4.0 (http://creativecommons.org/licenses/by-sa/4.0)], via Wikimedia Commons*

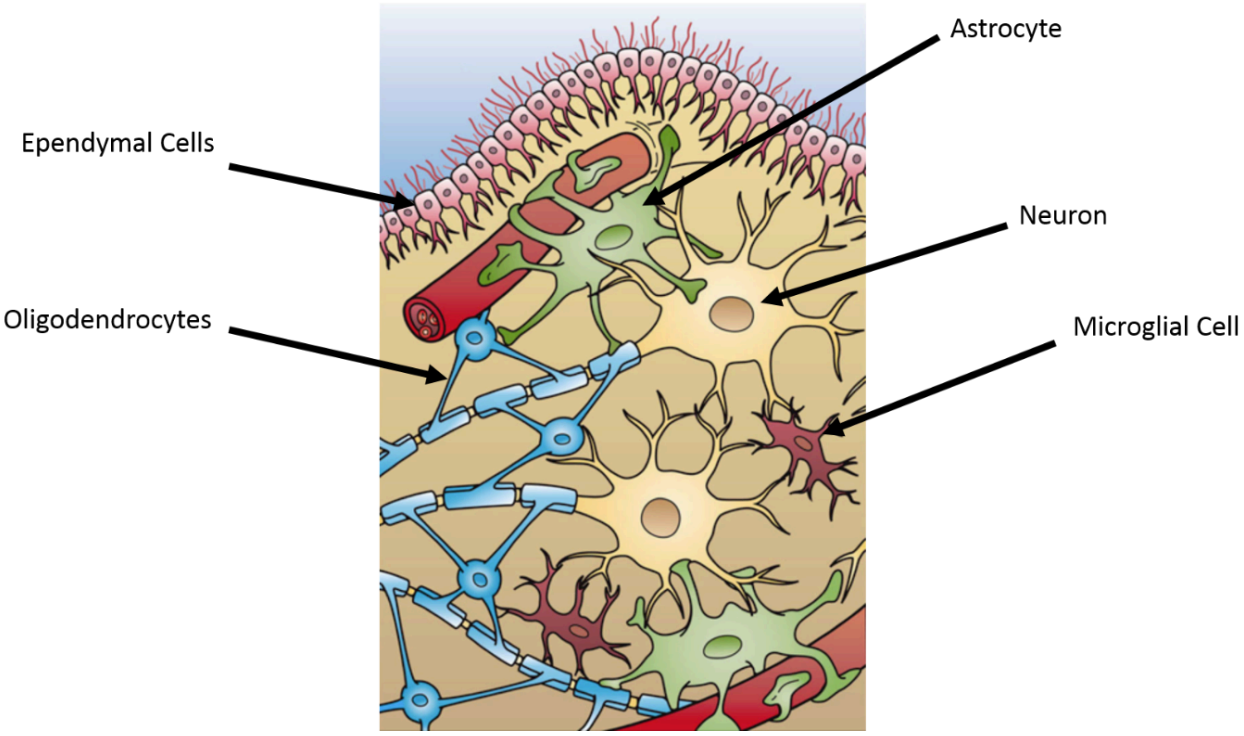
There are many diseases that selectively damage or destroy myelin; the most common demyelinating disease of the CNS is **multiple sclerosis**. Multiple sclerosis (MS) is an autoimmune disease that results in the selective destruction of oligodendrocytes, resulting in a reduction of myelin. The reduction in myelin severely decreases the conduction velocity and duration of action potentials in the affected neuron. Depending of where in the CNS the affected nurons are found, this can result in loss of sensory perception and motor control. The cause of MS is currently unknown, but the disease is twice as common in women as in men.

Ependymal Cell

Ependymal cells line the cavities (ventricles) of the CNS. Ependymal cells are responsible for the production of Cerebral Spinal Fluid (CSF) and are important barriers between the cerebral spinal fluid and the brain extracellular space. These cells beat their cilia to help circulate the cerebral spinal fluid. Ependymal cells selectively utilized different components of blood to create the CSF.

The Microglial Cell

Microglial cells are immune cells that are rapidly activated in the CNS in response to injury or infection. . Injury causes these cells to proliferate, change shape, and become phagocytic. These cells are also very important in presenting antigens to lymphocytes in response to infection. Although these cells are an important component of the CNS, it is believed that their chronic activity is also toxic to neurons and can result in long-term damage. For this reason, medical intervention in response to brain injury often involves factors that inhibit microglial activity.



Glial Cells of the CNS.

Derived from Wikimedia Commons Dec 2013; Author: Holly Fischer; License: Creative Commons Attribution 3.0 Unported license.

6.1.3

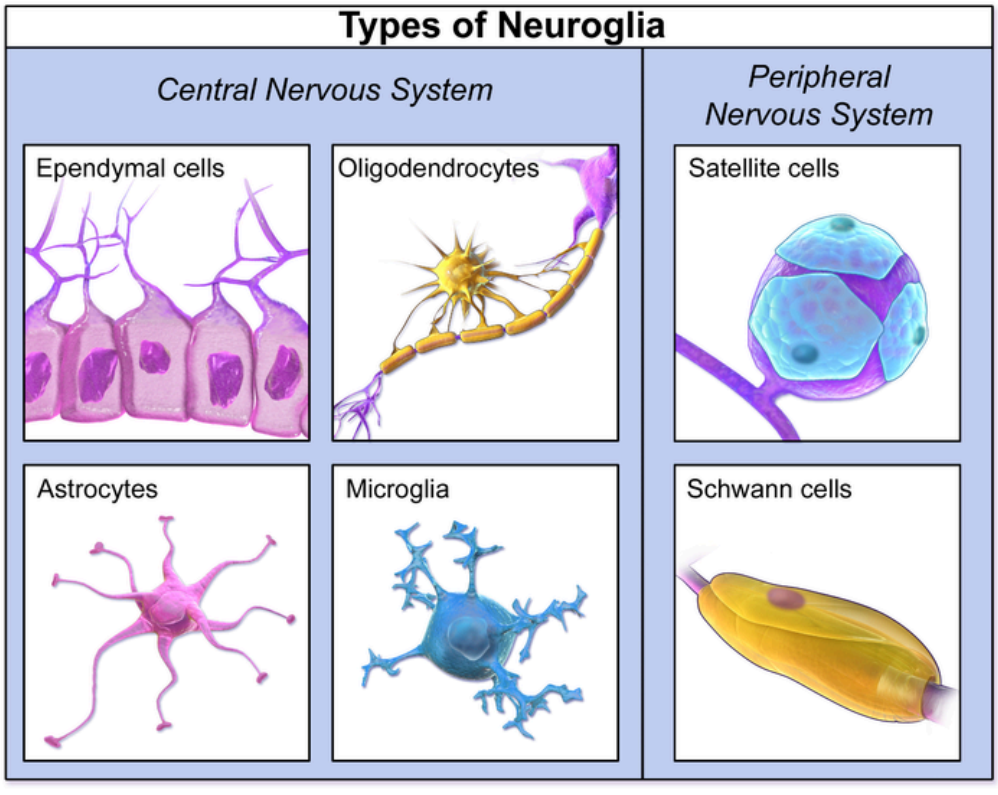
Glial Cells of the PNS

The Schwann Cell

The Schwann cell is the myelinating cell of the PNS. In contrast to the oligodendrocyte of the CNS, which uses multiple processes to myelinate multiple segments of axons, a Schwann cell provides myelin for a single segment of an axon. Still, the appearance and function of myelin in the PNS is exactly the same as the CNS.

The Satellite Cell

Satellite glial cells help regulate the external chemical environment around neurons of the PNS. In this way, they are very similar to the astrocyte of the CNS but, in addition, are highly sensitive to injury and inflammation.



Types of Neuroglia. Author:Blausen.com staff (2014).
"Medical gallery of Blausen Medical 2014". WikiJournal of Medicine 1 (2). DOI:10.15347/wjm/2014.010. ISSN 2002-4436.

6.2

PHYSIOLOGY OF THE NEURON

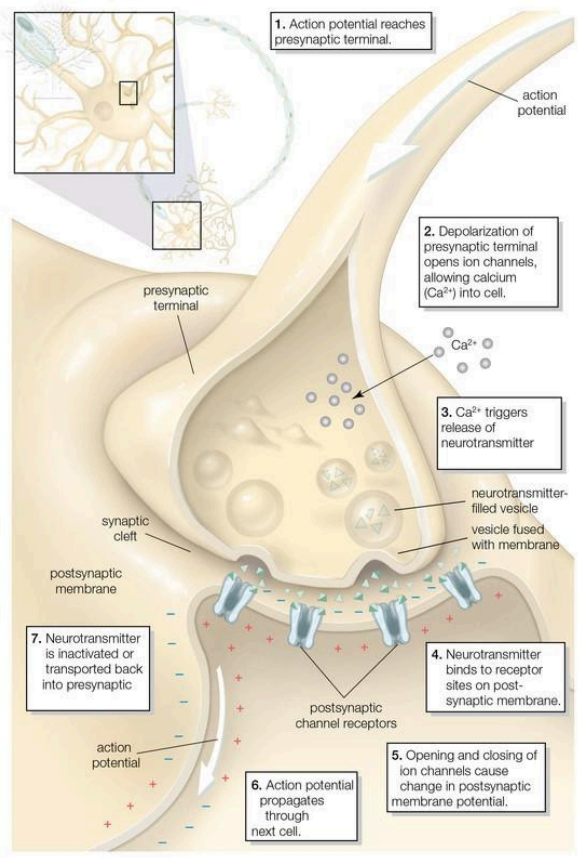
Typically, voltage changes in neurons flow from dendrites, to the soma, and to the axon. In sensory neurons, however, environmental stimuli (light, chemicals, and pain) activate ion channels which produce action potentials that flow from the axon to the soma. In either case, neurons propagate signals along their axons in the form of action potentials, which is how neurons communicate with other neurons or cells. The communication that occurs between these cells is called **synaptic transmission**.

6.2.1

The Synapse

Structurally, two types of synapses are found in neurons: chemical and electrical. **Chemical synapses** occur when neural membranes are very close together but remain distinct, leaving a space called the synaspe. **Electrical synapses** occur when membranes are linked together (gap junctions) via specialized proteins (connexins) that allow the flow of ions quickly from one cell to another. Electrical synapses are found in heart muscle and in various other cells. Because electrical synapses are rare in the nervous system, the remainder of this section will address the chemical synapse.

Chemical synapses use chemicals called *neurotransmitters* to communicate the messages between cells. The part of the synapse that releases the neurotransmitter into the synapse is called the *presynaptic terminal*, and the part of the synapse that receives the neurotransmitter is called the *postsynaptic terminal*. The narrow space between the two regions is called the *synaptic cleft*. Both the presynaptic and postsynaptic terminals contain the molecular machinery needed to carry out the signaling process. The presynaptic terminal contains large numbers of vesicles that are packed with neurotransmitters. When an action potential arrives at the presynaptic terminal, voltage-gated Ca^{++} channels open, which allows for the influx of Ca^{++} which then activates an array of molecules called SNARE proteins in the neuronal membrane and the vesicular membrane. These newly activated SNARE proteins cause the vesicle containing neurotransmitter to fuse with the presynaptic terminal membrane which leads to exocytosis of the vesicles, which results in the release of the neurotransmitter. The neurotransmitter then diffuses across the synaptic cleft and binds to receptors located in the postsynaptic membrane and induces a conformational change. These ligand gated channels undergo a conformation change which causes the receptor to act as a pore in the membrane for ions to move through. Depending on the type of ion, the effect on the postsynaptic cell may be depolarizing (excitatory) or hyperpolarizing (inhibitory). To turn off the signal there are enzymes that reside in the synaptic cleft that breakdown and inactivate the neurotransmitters. The components of the neurotransmitter are then taken back up by the presynaptic terminal to be recycled to make more of the neurotransmitter. An example of one of the enzymes is acetylcholinesterase that breaks down the neurotransmitter acetylcholine.



Synapse *© 2013 Encyclopædia Britannica, Inc. Taken from BYUI Image Quest Dec 2013.*

If the neurotransmitter causes the membrane post synaptic potential to go towards threshold it is called an **EPSP** which is the abbreviation for an *excitatory post synaptic potential*, whereas an inhibitory response takes the membrane potential away from threshold (further towards hyperpolarization) and is called an **IPSP** or *inhibitory post synaptic potential*.

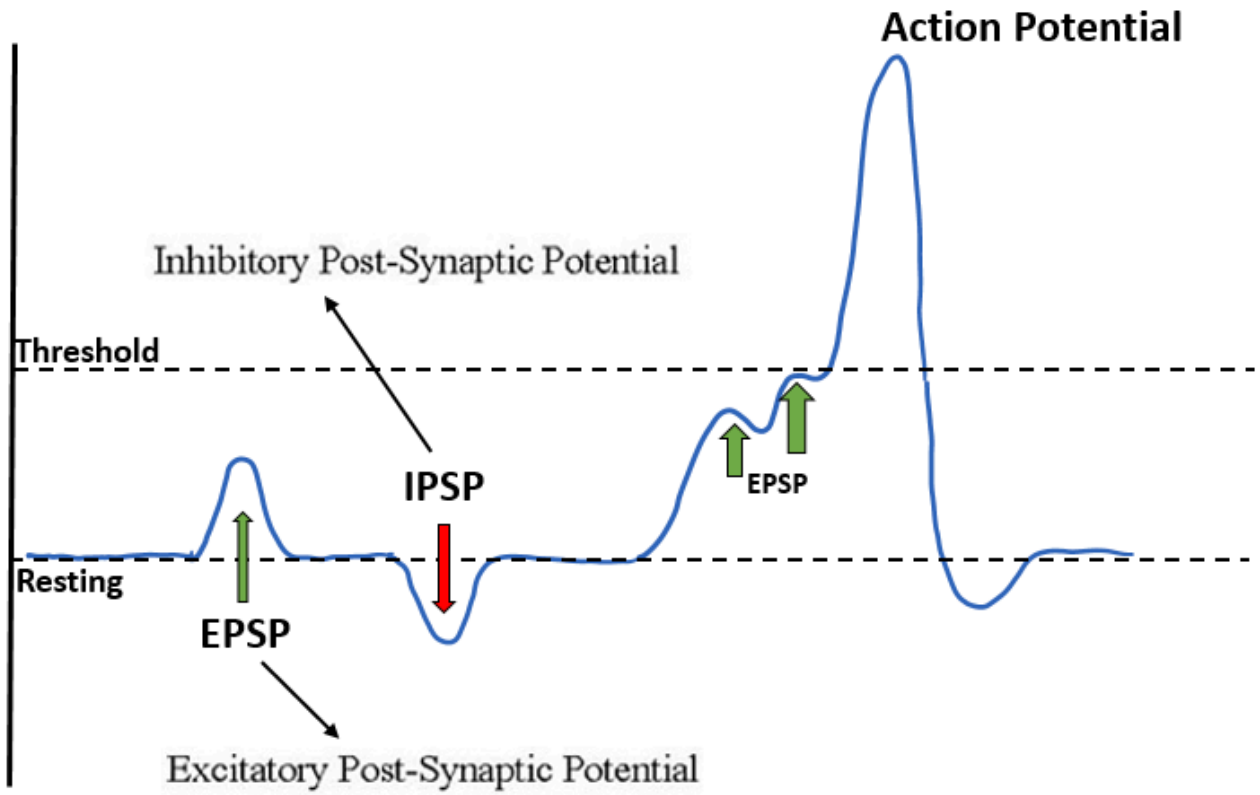


Image by BYU-I student 2017

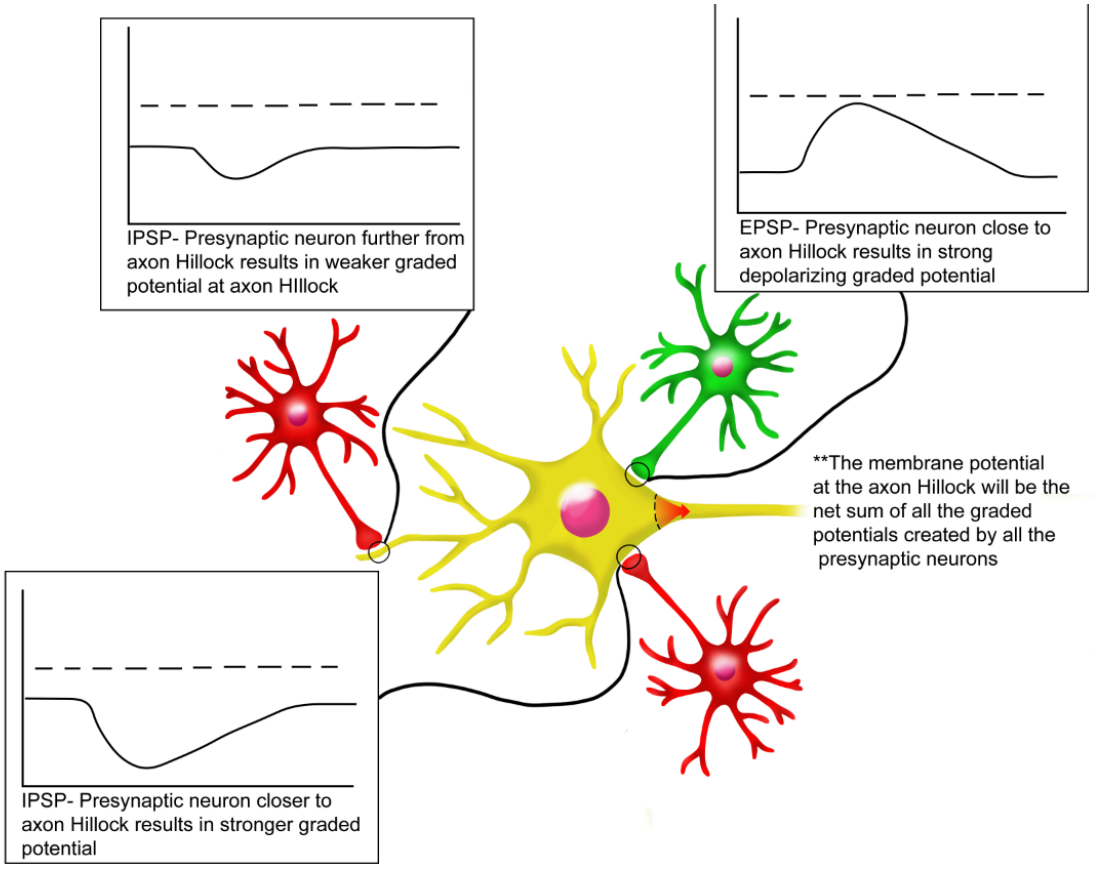


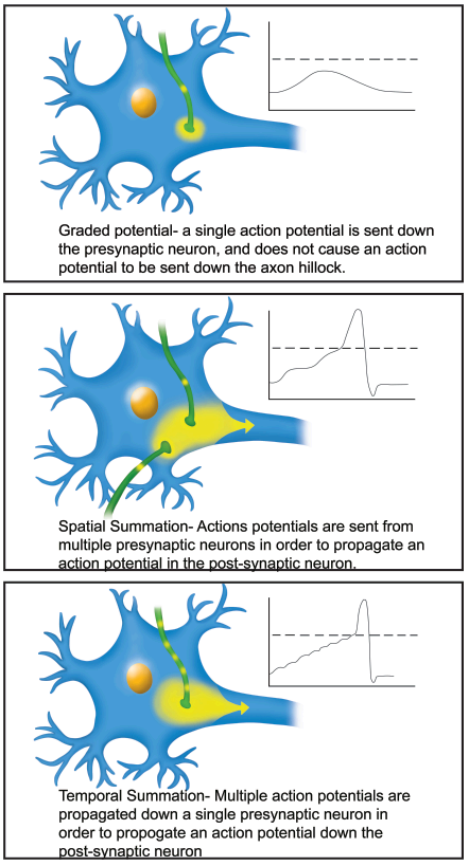
Image by Becky T. BYU-Idaho, 2018.

A cell body will have many synapses on it and on its surrounding dendrites. Some of the synapses will result in the cell body membrane potential moving closer to threshold. Other synapses result in the cell body membrane potential moving farther from threshold (hyperpolarization). As mentioned, any synapse that moves the potential closer to threshold is called an *excitatory post synaptic potential*, and any synapse that moves the potential farther from threshold is called an *inhibitory post synaptic potential*. The net effect of all the EPSPs and IPSPs is experienced at the axon hillock. If threshold is reached at the axon hillock, then an action potential will continue down the axon.

The ultimate goal of an EPSP is to cause enough change in the membrane to initiate an action potential. The goal of the IPSP is to cause a change in the membrane to prevent an action potential. Each EPSP or IPSP lasts a few milliseconds, and then, the membrane returns to the original resting membrane potential. In many cases, a single EPSP is not sufficient to cause an action potential. Therefore, many EPSPs from multiple synapses can combine at the axon hillock, which results in a much larger voltage change that can exceed threshold and cause an action potential. This phenomenon is called **spatial summation**. EPSPs from the same synapse can also combine if they arrive in rapid succession; this phenomenon is called **temporal summation**. Requiring multiple EPSPs to fire an action potential is a way that neurons increase sensitivity and accuracy.

6.2.2

Summation



This Image shows temporal and spatial summation using EPSP synapses as an example. IPSP synapses can also occur in both temporal and spatial summation. The difference would be that IPSP synapses try to drive the membrane potential further down and away from threshold. Whether the membrane potential measured at the axon hillock reaches threshold or not depends on the net effect of all the EPSP and IPSP summations. *Image by BYU-I student - Becky T. 2018*

A response as an EPSP or an IPSP will depend on the type of neurotransmitter/receptor combination present in the synapse. There are over a hundred known neurotransmitters, and many of them have unique receptors. Receptors can be divided into two broad groups: chemically-gated ion channels (ionotropic) and second messenger systems (metabotropic). When chemically (ligand) gated ion channels are activated, certain ions are allowed to flow across the membrane. The ion type will determine whether the result is an EPSP or an IPSP. When a second messenger system is activated, it results in a cascade of molecular interactions within the target or postsynaptic cell. The type of cascade that is elicited will result in the response being either excitatory or inhibitory.

Excitatory Synapses

Most excitatory synapses in the brain use glutamate or aspartate as the neurotransmitter. These neurotransmitters bind to non-selective cationic channels that allow for Na^+ and K^+ to pass. As mentioned earlier, it takes many EPSPs from these kinds of synapses to depolarize a postsynaptic neuron enough to reach threshold of the axon hillock and trigger an action potential.

A very important subset of synapses in the brain includes a group capable of forming memories by increasing the activity and the strength of the synapse. This process is called **long-term potentiation**. Long-term potentiation operates at the synapse, using the neurotransmitter glutamate and the receptor known as the NMDA receptor. The NMDA receptor is unique in that it is both ligand and voltage regulated. When activated by ligands, it becomes permeable to Na^+, but if the charge difference is sufficient, the channel becomes permeable to Ca^{++} as well. Ca^{++} can initiate a second messenger cascade that results in an increase in the number of glutamate receptors, thereby increasing the strength of the synapse, making it easier to reach threshold in the post-synaptic cell. The change in strength can last for weeks, months, or even years depending on whether or not the synapse is continually used. Long-term potenitiaion is a major mechanism underlying memory and learning.

Inhibitory Synapses

It may seem somewhat of a paradox to have inhibitory synapses, but the excitability of neurons is essentially governed by a balance between excitation and inhibition. The main inhibitory neurotransmitters are GABA and glycine. Both neurotransmitters bind to receptors that result in an increase conductance of Cl^-. Because of the negative charge of Cl^- and the fact that it usually moves into the cell, the effect is to oppose depolarization and cause the membrane to move away from threshold (IPSP).

7.0

MODULE 7: SKELETAL MUSCLE

7.1

FUNCTIONS AND PROPERTIES OF SKELETAL MUSCLE TISSUE

1. **Movement:** Our body's skeleton gives enough rigidity to our body so that *skeletal muscles* can yank and pull on it resulting in body movements such as walking, chewing, running, lifting, and manipulating objects with our hands.
2. **Maintenance of Posture:** Without much conscious control, our muscles generate a constant contractile force that allows us to maintain an erect or seated position, also known as *posture*.
3. **Respiration:** Our muscular system automatically drives movement of air into and out of our body.
4. **Heat Generation:** Contraction of muscle tissue generates heat, which is essential for maintenance of temperature homeostasis. For instance, if our core body temperature falls we shiver to generate more heat.
5. **Communication:** Muscle tissue allows us to talk, gesture, write, and convey our emotional state by doing such things as smiling or frowning.

All muscle cells share several properties: contractility, excitability, extensibility, and elasticity.

1. **Contractility** is the ability of muscle cells to forcefully shorten. For instance, in order to *flex* (decrease the angle of a joint) your elbow, you need to *contract* (shorten) the biceps brachii and other elbow flexor muscles in the anterior arm. Notice that in order to *extend* your elbow, the posterior arm extensor muscles need to contract. *Thus, muscles can only pull, never push*.
2. **Excitability** is the ability to respond to a stimulus, which is delivered from a motor neuron.
3. **Extensibility** is the ability of a muscle to be stretched. For instance, let's reconsider our elbow flexing motion we discussed earlier. In order to be able to flex the elbow, the elbow extensor muscles must extend in order to allow flexion to occur. Lack of extensibility is known as *spasticity*.
4. **Elasticity** is the ability to recoil or bounce back to the muscle's original length after being stretched.

Skeletal muscle is also known as **voluntary muscle** because we can consciously, or voluntarily, control it in response to input by nerve cells. Skeletal muscle is also referred to as **striated** ("striped") because it has a microscopically streaked or striped appearance. Skeletal muscle and its associated connective tissue comprise about 40% of our weight. Skeletal muscle also has a unique characteristic with regard to nuclei. There are many nuceli in each skeletal muscle cell. These nuclei are generally pressed up against the cell membrane as there is very little room inside the cells given all the contractile proteins that are there.

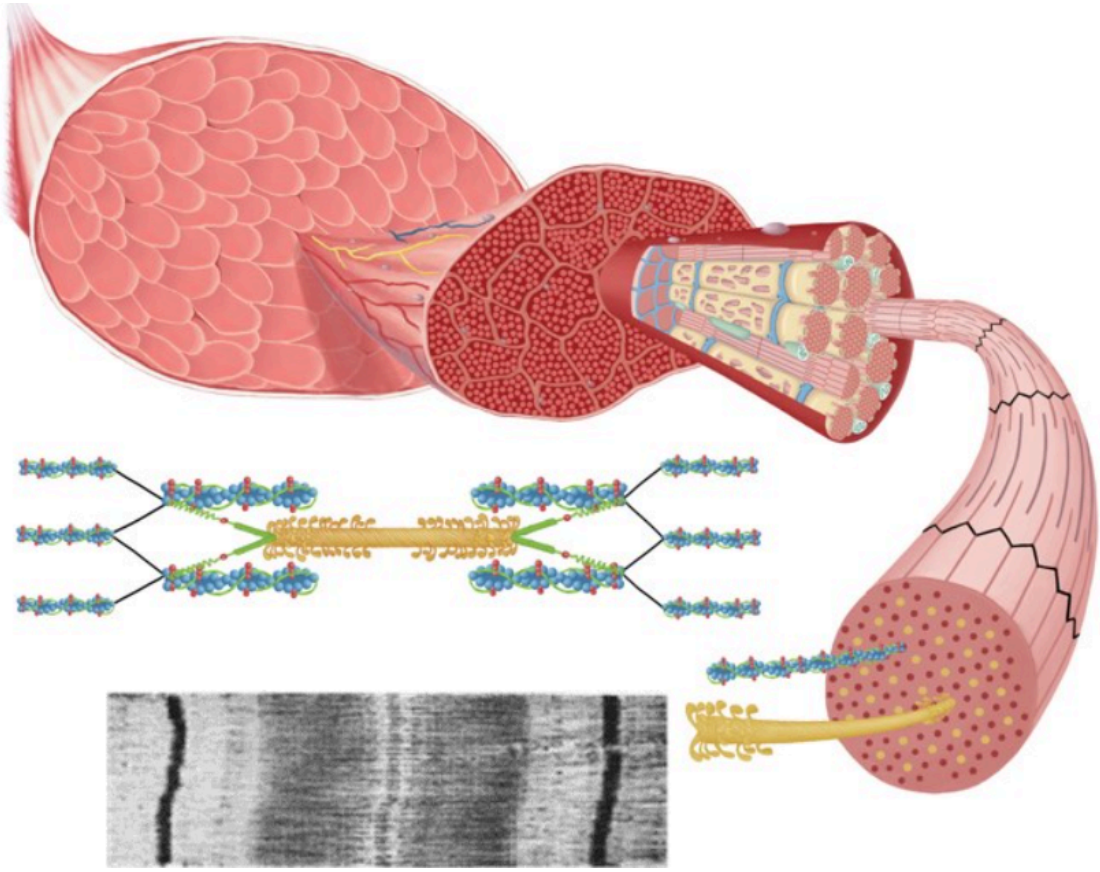
7.2

SKELETAL MUSCLE ORGANIZATION

Before completing this reading, you should complete the following:

Step 01

Download a page sized image of "Muscle Organization" shown below



Skeletal Muscle Organization. *Image drawn by BYU-Idaho student Nate Shoemaker Spring 2016*

Step 02

Work through a tutorial that will help you label your image.

Step 03

With your newly labeled image in hand, read through the following paragraphs.

7.2.1

Gross and Microscopic Structure

Each skeletal muscle cell, also called a **muscle fiber**, develops from many individual embryonic **myocytes** that fuse into one long multi–nucleated skeletal muscle cell. The resultant fused muscle fiber is the length of the entire muscle. The longest muscle fiber in the body belongs to the sartorious muscle on the leg. These muscle fibers are bound together into bundles, or **fascicles**, and are supplied with a rich network of blood vessels and nerves. The fascicles are then bundled together to form the intact muscle. Muscle fibers are the same diameter as hair follicles. As you look down at your bicep, visualize small strands of hair follicles extending from your shoulder to the radius bone in your forearm. Clearly, one hair follicle would be too fragile to move your arm, but hundreds of millions are very adequate! Let's dissect a skeletal muscle, beginning with the muscle as a whole externally and continuing internally down to the submicroscopic level of a single muscle cell. In an intact skeletal muscle, a rich network of nerves and blood vessels nourish and control each muscle cell. These muscle fibers are individually wrapped and then bound together by several different layers of fibrous connective tissue.

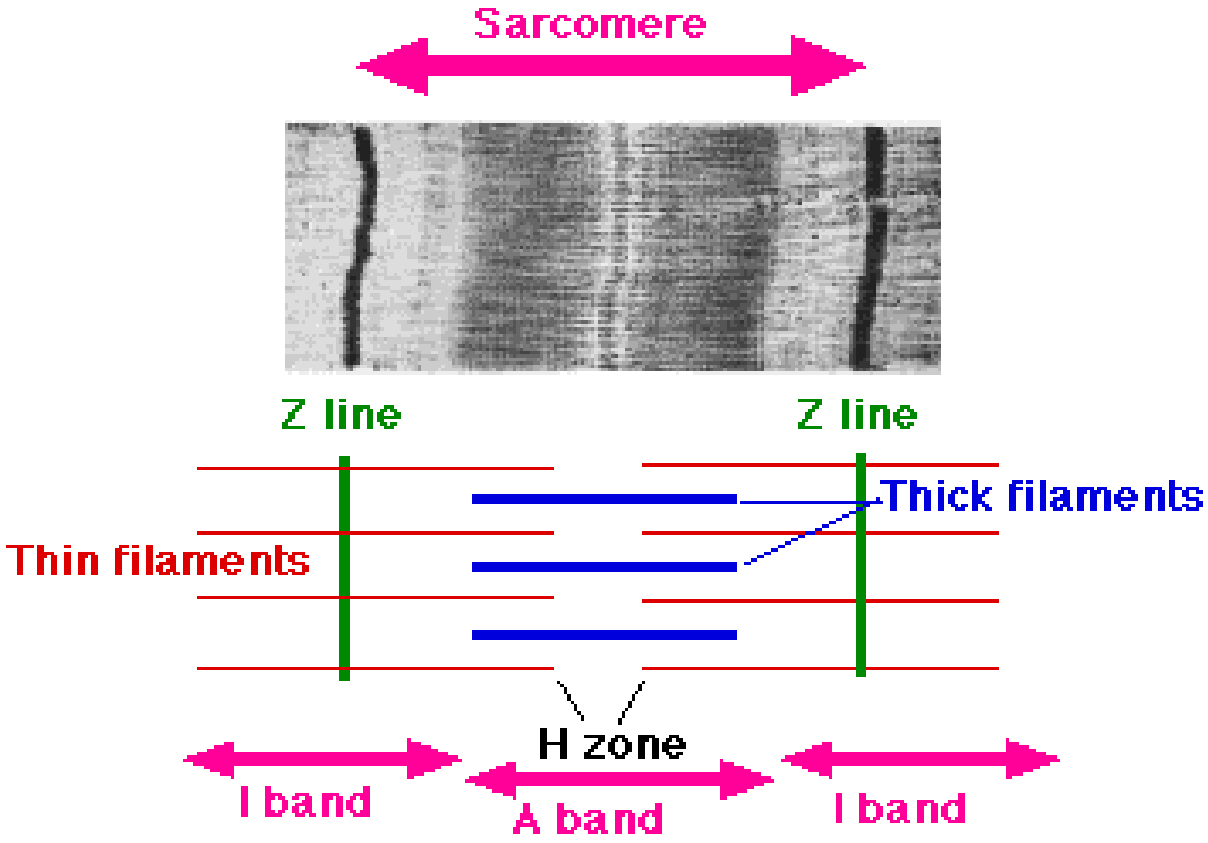
The **epimysium** (*epi* means “outside,” and mysium means “muscle”) is a layer of dense fibrous connective tissue that surrounds the entire muscle. This layer is also often referred to as the **fascia**. Each skeletal muscle is formed from several bundled fascicles of skeletal muscle fibers, and each fascicle is surrounded by **perimysium** (*peri* means “around”). Each single muscle cell is wrapped individually with a fine layer of loose (areolar) connective tissue called **endomysium** (*endo* means “inside”). These connective tissue layers are continuous with each other, and they all extend beyond the ends of the muscle fibers themselves, forming the **tendons** that anchor muscles to bone, moving the bones when the muscles contract.

Deep to the endomysium, each skeletal muscle cell is surrounded by a cell membrane known as the **sarcolemma** (you will see the prefixes *sarc-* and *myo-* quite a bit in this discussion, so you should understand that these are prefixes that refer to "muscle"). The cytoplasm, or **sarcoplasm**, contains a large amount of *glycogen* (the storage form of glucose) for energy, and **myoglobin** (a red pigment similar to hemoglobin that can store oxygen). Most of the intracellular space, however, is taken up by cylindrical (rod-like) **myofibril** protein structures. Each muscle fiber contains hundreds or even thousands of myofibrils that extend from one end of each muscle fiber to the other. These myofibrils take up about 80% of the intracellular space and are so densely packed inside these cells that mitochondria and other organelles get sandwiched between them while the nuclei get pushed to the outside and are located on the periphery, right under the sarcolemma.

Each myofibril is comprised of several varieties of protein molecules that form the **myofilaments**, and each myofilament contains the contractile segments that allow contraction. These contractile segments are known as **sarcomeres** (*sarc-* means “muscle,” and *mere* means “part”). The striations seen microscopically within skeletal muscle fibers are formed by the regular, organized arrangement of myofilaments—much like what we would see if we painted stripes on chopsticks and bundled them together with plastic wrap, with the plastic wrap representing the sarcolemma.

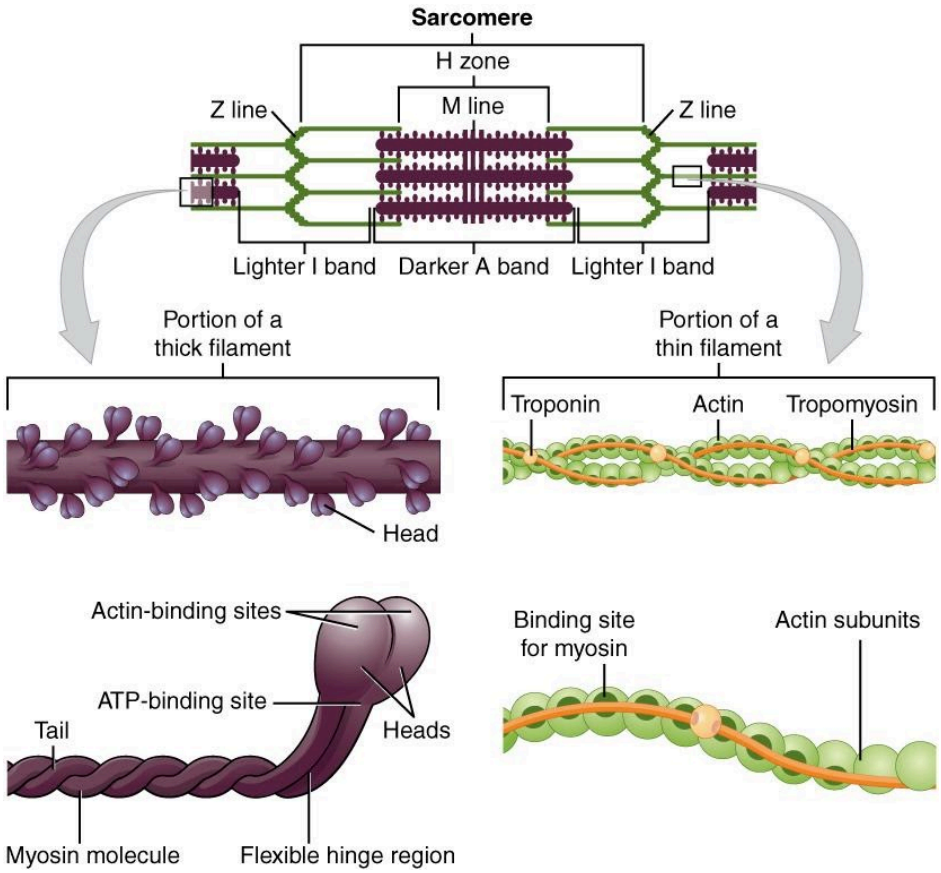
The striations microscopically visible in skeletal muscle are formed by the regular arrangement of proteins inside the cells. Notice that there are light and dark striations in each cell. The dark areas are called **A bands**, which is fairly easy to remember because "A" is the second letter in "dark." The light areas are called **I bands** and are also easy to remember because "i" is the second letter in "light." ("A" actually stands for *anisotropic*, and "I" stands for *isotropic*. Both of these

terms refer to the light absorbing character of each band. However, we'll stick to A and I bands.) The image below shows a micrograph of a sarcomere, along with a drawing representing the different parts of the sarcomere.



Skeletal Muscle Sarcomere: Thick and Thin Filaments, Z Line, H Zone, I & A Bands. *File:Sarcomere.gif; Author: Sameerb; Site: https://commons.wikimedia.org/wiki/File:Sarcomere.gif; License: Public Domain, No restrictions*

Notice that in the middle of each I band is a darker line called the **Z line** or **Z disc**. The Z lines are the divisions between the adjacent sarcomeres. Sarcomeres are connected, end to end, along the entire length of the myofibril. Also, in the middle of each A band is a lighter **H zone** (H for *helle*, which means "bright"), and each H zone has a darker **M line** (M for "middle") running right down the middle of the A band.



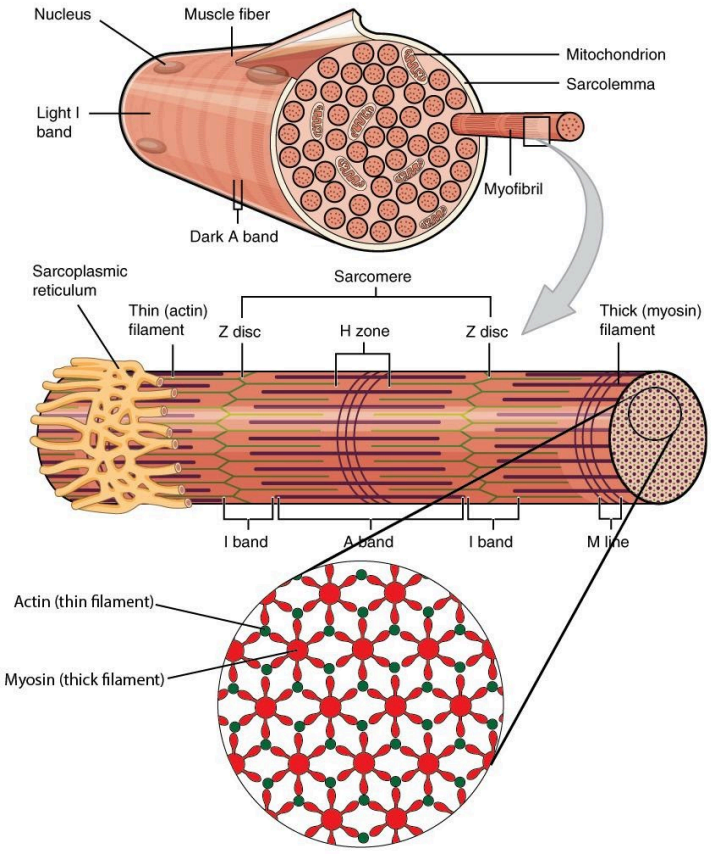
Sarcomere: Detailed Illustration of Thick and Thin Filaments: Title: 1003_Thick_and_Thin_Filaments.jpg; Author: OpenStax College; Site: https://books.byui.edu/-fuq;
License: licensed under a Creative Commons Attribution 4.0 License.

Each myofibril, in turn, contains several varieties of protein molecules, called **myofilaments**. The larger myofilaments, known as **thick myofilaments**, are made of the protein **myosin**, and the smaller **thin myofilaments** are chiefly made of the protein **actin**.

Let's discuss each myofilament in turn. Each actin molecule is composed of two strands of *fibrous actin (F-actin)* and a series of **troponin** and **tropomyosin** molecules. Each F-actin is formed by two strings of *globular actin (G-actin)* wound together in a double helical structure, much like twisting two strands of pearls with each other. Each G-actin molecule would be represented by a pearl on our hypothetical necklace. Each G-actin subunit has a binding site for the myosin head to attach to the actin. **Tropomyosin** is a long string-like polypeptide that parallels each F-actin strand and functions to either hide or expose the "active sites" on each globular actin molecule. Each tropomyosin molecule is long enough to cover the active binding sites on seven G-actin molecules. These proteins run, end to end, along the entire length of the F-actin. Associated with each tropomyosin molecule is a third polypeptide complex known as **troponin**. Troponin complexes contain three globular polypeptides (*Troponin I, Troponin T,* and *Troponin C*) that have distinct functions. Troponin I binds to actin, troponin T binds to tropomyosin and helps position it on the F-actin strands, and troponin C binds calcium ions. There is one troponin complex for each tropomyosin. When calcium binds to troponin C, it causes a conformational change in the entire complex that results in exposure of the myosin binding sites on the G-actin subunits. More on this later.

The thick myofilaments are composed chiefly of the protein **myosin**, and each thick myofilament is composed of about 300 myosin molecules bound together. Each myosin is made up of six protein subunits, two *heavy chains* and four *light chains*. The heavy chains have a shape similar to a golf club, having a long shaft-like structure, to which is connected the globular myosin head. The shafts, or tails, wrap around each other and interact with the tails of other myosin molecules, forming the shaft of the thick filament. The globular heads project out at right angles to the shaft. Half of the

myosin molecules have their heads oriented toward one end of the thick filament, and the other half are oriented in the opposite direction. It is the myosin heads that bind to the active sites on the actin. The connection between the head and the shaft of the myosin molecules function as a hinge and as such is referred to as the **hinge region**. The hinge region can bend and, as we shall see later, creates the power stroke when the muscle contracts. The center of the thick filament is composed only of the shaft portions of the heavy chains. Additionally, each myosin head has an ATPase that binds to and hydrolyzes ATP during muscle contraction. It is the ATP that provides the energy for muscle contraction. Each of the myosin heads is associated with two myosin light chains that play a role in regulating the actions of the myosin heads, but the exact mechanism is not fully understood. The three-dimensional arrangement of the myosin heads is very important. Imagine that you were looking at a thick filament from the end, and there is a myosin head sticking straight up. As you moved around the circumference of the thick filament, you would see myosin heads every 60 degrees. This allows each thick filament to interact with six thin filaments. Likewise, each thin filament can interact with three thick filaments. This arrangement requires that there be two thin filaments for every thick filament in the myofibril (see image below).



Muscle Fiber Detailed Diagram. Adapted from the following image: Title: 1022_Muscle_Fibers_(small).jpg;Author: OpenStax College; Site: https://books.byui.edu/-fuq;License: licensed under a Creative Commons Attribution 4.0 License.

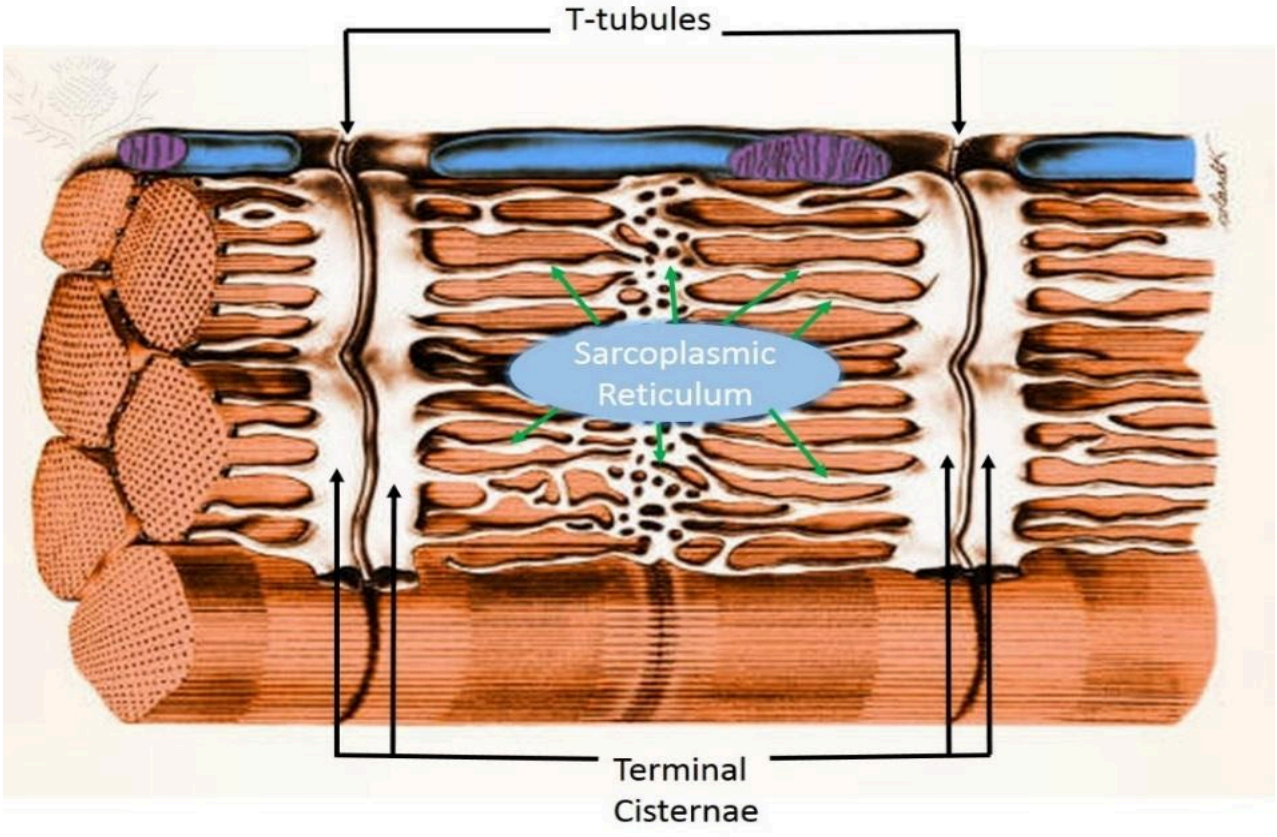
During muscle contraction, the myosin heads link the thick and thin myofilaments together, forming **cross bridges** that cause the thick and thin myofilaments to slide over each other, resulting in a shortening of each sarcomere, each skeletal muscle fiber, and the muscle as a whole—much like the two parts of an extension ladder that slide over each other. To summarize, in order for the shortening of the muscle to occur, the myosin heads have three important properties: 1.) The heads can bind to active sites on G-actin molecules, forming *cross bridges*. 2.) The heads are attached to the rod-like portions of the heavy myosin molecules by a *hinge region* as already discussed. 3.) The heads have *ATPase* enzymes that can break down ATP, using the resulting energy to bend the hinge region and allow detachment of the myosin heads from actin.

The Z-line (or Z-disc) is composed of proteins (**alpha actinin**) which provide an attachment site for the thin filaments. Likewise, the M-line is composed of proteins (**myomesin**) that hold myosin molecules in place and creatine kinase enzymes (explained later). The A band is formed by myosin molecules, and the I band is the location where thin filaments do not overlap the thick filaments. The H zone is that portion of the A band where the thick and thin filaments do not overlap.

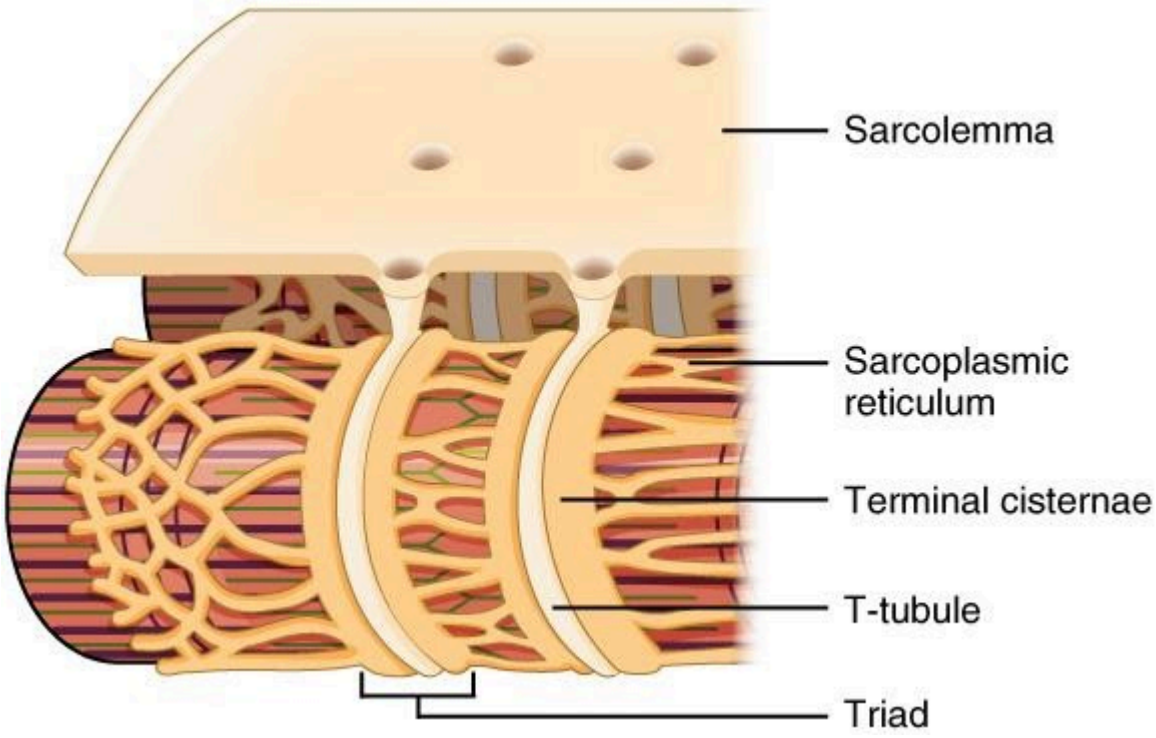
There is another important structural protein that extends from the Z disc to the M line, running within the thick filament. Due to its large size, this protein is called **titin** (titin is the largest known protein in the human body and has roughly 30,000 amino acids). It forms the core of the thick myofilaments, holding it in place, and thus keeps the A band organized. In addition, titin has the ability to stretch and recoil. It functions to prevent overstretching and damage to the muscle and to return the muscle to its normal length when it is stretched beyond its normal resting length. Recall that one of the properties of muscle is its elasticity. Titin is the protein responsible for this property.

There are several other important structural proteins, but we will only discuss one more: **dystrophin**. Dystrophin is a protein located between the sarcolemma and the outermost myofilaments. It links actin to an integral membrane protein, which, in turn, links the muscle cell to the endomysium of the entire muscle fiber. Without dystrophin the muscle will be able to contract (shorten) but no force (ie., limb movement) can be generated. Genetic mutation of the gene coding for dystrophin is one of the root causes of a class of muscle diseases known collectively as *muscular dystrophy (MD)*. The most common form of MD is Duchene muscular dystrophy (DMD), which is inherited in a "sex-linked" fashion and affects boys. Most DMD patients become wheelchair bound early in life, usually by age 12 or so. Difficulty breathing usually becomes problematic by age 20 and sadly is often the cause of their premature death.

Sarcoplasmic Reticulum and T Tubules



Sarcoplasmic Reticulum. *© 2013 Encyclopædia Britannica, Inc. Downloaded and labeled from image quest BYU-Idaho 2013.*



T-Tubule. *Title: 1023_T-tubule.jpg; Author: OpenStax College; Site: http://cnx.org/contents/6df8aab3-1741-4016-b5a9-ac51b52fade0@3/Skeletal-Muscle; License: licensed under a Creative Commons Attribution 4.0 License.*
There are two sets of tubules within skeletal muscles fibers that carry out critical functions during muscle contractions: the sarcoplasmic reticulum and the T-tubules.

T-tubules (transverse tubules) are invaginations, or indentations, of the sarcolemma. They are formed much like poking holes with a fork in your raw potato before cooking it. T-tubules communicate with the extracellular space and are filled with extracellular fluid. They are located on the sarcomere at the point where the A band and I band overlap. The T-tubules are flanked on either side by dilated regions of the cell's endoplasmic reticulum—the sarcoplasmic reticulum.

Sarcoplasmic reticulum (SR) is an elaborate network of smooth endoplasmic reticulum that surrounds and encases each myofibril, much like a loosely knitted sweater that covers your arms. It stores calcium which can then be released into the sarcoplasm when an action potential is conducted along the sarcolemma of the T-tubule. Most of the sarcoplasmic reticulum runs parallel to the myofibrils, but there are right-angle enlargements of the SR at the A band/I band junctions that flank the T-tubules. These enlargements are known as **terminal cisternae** ("end sacs") (see the image above). One T-tubule along the` two terminal cisternae that parallel it form the **triad**. The triad is critical in skeletal muscle function. At each triad, the T-tubule membrane contains large numbers of voltage-dependent proteins called dihydropyridine (DHP) channels or L-type calcium channels. Although these are called channels, they do not allow calcium to move through them; rather, they are physically linked to calcium release channels on the terminal cisternae known as ryanodine receptor channels (RyR). When the membrane is depolarized by an action potential, the DHP channel detects a depolarization and causes the RyR channels to open, resulting in the release of calcium from the terminal cisternae of the SR.

7.3

NEUROMUSCULAR JUNCTION, EXCITATION-CONTRACTION COUPLING, SLIDING FILAMENT THEORY, CONTRACTURES AND CRAMPS

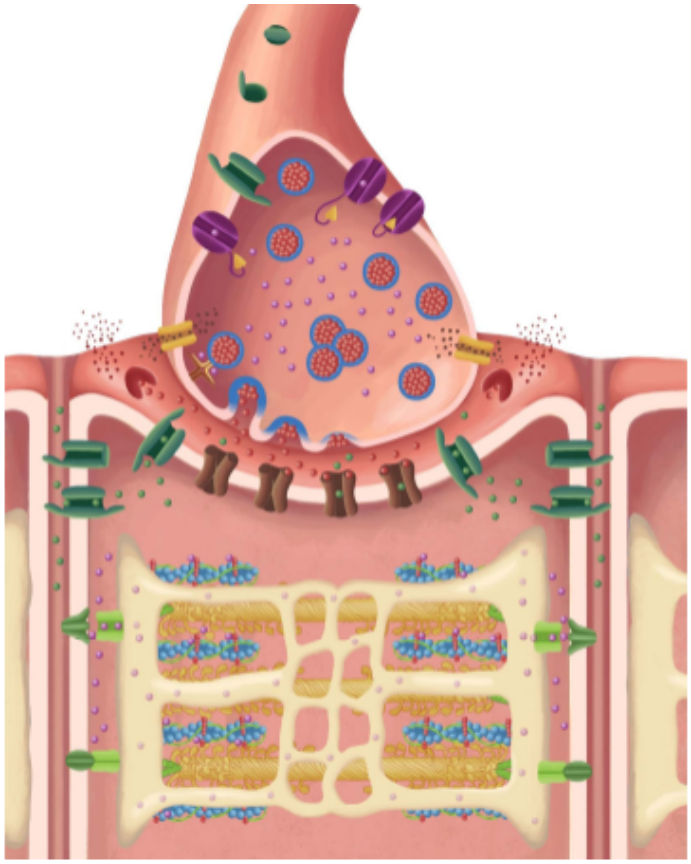
7.3.1

Neuromuscular Junction, Excitation-Contraction Coupling, and Sliding Filament Theory

An important part of understanding the full story of muscle contraction is understanding how a nerve communicates an electrical signal to a muscle fiber. To help with this, you should complete the following steps before reading further.

Step 01

Download a page sized image of the neuromuscular junction.



Neuromuscular Junction. Image drawn by BYU-Idaho student Spring 2013

Step 02

Complete an image labeling tutorial and label your own blank image.

Step 03

With your newly labeled image in hand, finish the rest of the reading on this page.

In order for these muscle fibers to contract, there needs to be an *electrical event* (an *action potential*) that is followed by a *mechanical event* (the contraction of the muscle fiber). Because you have already learned about the resting membrane potential and the action potential, we'll move past that and talk about the **sliding filament model of muscle contraction.**

Recall that we have already mentioned the fact that the thick and thin myofilaments slide over each other, like the parts of an extension ladder. The proteins themselves don't shorten. The muscle contraction and shortening occur as the myofilaments grip each other, slide past each other, and shorten the sarcomeres. Thus, this is known as the *sliding filament model of muscle* contraction. Let's also remember that in order for action potentials to both start and propagate (travel), it is necessary for various *ion* channels to open and close at just the right time. Some of these ion channels open in response to the binding of a *ligand*—an atom or molecule that binds to a receptor and stimulates a specific response. These types of ion channels are known as *ligand-gated ion channels.* Other ion channels open in response to a change in *voltage (*electricity) and are known as *voltage-gated ion channels*, and others respond because they are mechanically linked to another channel and are knows as *mechanically-gated ion channels.* Now, we'll discuss the sequence of events that occur when an action potential reaches the end of the motor neuron.

1. An action potential arrives at the axon terminal of a somatic motor neuron. The axon terminal of the motor neuron connects to the muscle fiber via the neuromuscular junction (a synapse).
2. The arrival of the action potential stimulates *voltage-gated Ca2+ channels* in the axon's membrane to open, and Ca2+ enters the axon terminal from the extracellular space. *Note: these channels are on the neuron, not the muscle, be careful not to mistake these channels for the *voltage-gated Ca2+ channels* or DHP found in the T-tubule of muscle cells.
3. The *axon terminal* contains *synaptic vesicles* filled with the neurotransmitter *acetylcholine (ACh)*. The increased Ca2+ levels is the signal that stimulates exocytosis of these synaptic vesicles and the release of ACh into the synaptic cleft.
4. The ACh diffuses across the synaptic cleft, binding to and activated ligand-gated ion channels (nicotinic type I) on the sarcolemma of the post synaptic tissue (the muscle fiber). This specialized region of the sarcolemma is known as the **motor end plate**, and this is the location of the ACh receptors (nicotinic type 1).
5. ACh binding causes the channel to open, hence the name *ligand-gated*. These ion channels are permeable to both Na+ and K+. However, more Na+ diffuses into the cell than K+ diffuses out of the cell. This may seem odd since both would be moving down their concentration gradients. The difference has to do with the charge on the membrane. Since the inside of the cell is negative, which will attract Na+, Na+ will be moving down **both** its concentration and its electrical gradients. Potassium, on the other hand, would move down its concentration gradient but ***against*** its electrical gradient (the negative charge inside the cell will attract the K+). The Na+ entering the cell *depolarizes* the sarcolemma, which then will cause the closely associated voltage-gated Na+ channels to open, initiating an action potential that spreads out from the neuromuscular junction. The action potential not only travels across the sarcolemma but also down the T-tubules. Remember, T-tubules are just invaginations (inward protrusions) of the sarcolemma and are filled with extracellular fluid that is high in sodium (Na+) and low in potassium (K+). Also, please notice that the ACh receptors are *ligand-gated,* but movement of Na+ through them causes the closely associated v*oltage-gated* Na+ channels to open, resulting in *generation* and *propagation* of an action potential.
    a. While the action potential spreads, let's take a break and describe how the stimulation of the ACh receptors is terminated. For the muscle to relax, ACh must be removed from the synaptic cleft. This is done when ACh is cleaved (split) by an enzyme that resides in the cleft called *acetylcholinesterase.* This enzyme splits ACh into its two components, *acetate (acetyl)* and *choline*, rendering it nonfunctional. The acetate portion of acetylcholine diffuses out of the synaptic cleft. The *choline*, which is an essential nutrient in the Vitamin B group (B4), is taken up by the axon terminal, where it is recycled to make more acetylcholine. Although our bodies can make choline, we cannot produce enough for our needs and must get it in our diet and recycle what we have.
6. The action potential does its thing (if you have forgotten the basics of action potentials, review module 5).
7. As the action potential spreads along the sarcolemma and the T-tubules, the resultant change in potential causes other voltage-gated channels in the T-tubule to respond. These channels are called dihydropyridine channels (DHP) or L-type Ca2+ channels and are mechanically linked to ryanodine receptor channels (RyR), which are calcium channels located in the sarcoplasmic reticulum membrane. These two protein channels span the distance between the T-tubule and the terminal cisternae of the sarcoplasmic reticulum. In response to the change in membrane potential, the DHP channel causes the RyR to open and allows Ca2+ ions to flow through it (RyR) from the sarcoplasmic reticulum into the sarcoplasm. These calcium ions bind to troponin, causing it to move the tropomyosin molecules off of the active sites on each G-actin molecule.
8. Uncovering the active sites allows the myosin heads to bind to the actin binding sites, forming **cross bridges**. In the resting state, the myosin head is "cocked" and ready to go. It also has ADP and phosphate (Pi) attached to it. The binding to actin releases energy, some of which is released as heat, and the remaining is capture in the phosphate bond which breaks the bond, releasing Pi and causing the myosin head to bend. This bending or **power stroke** forcefully pulls the actin past the myosin. During the power stroke, the ADP is also released from the myosin. Recall that in the arrangement of the thick filaments, half of the myosin molecules are pointing one way,

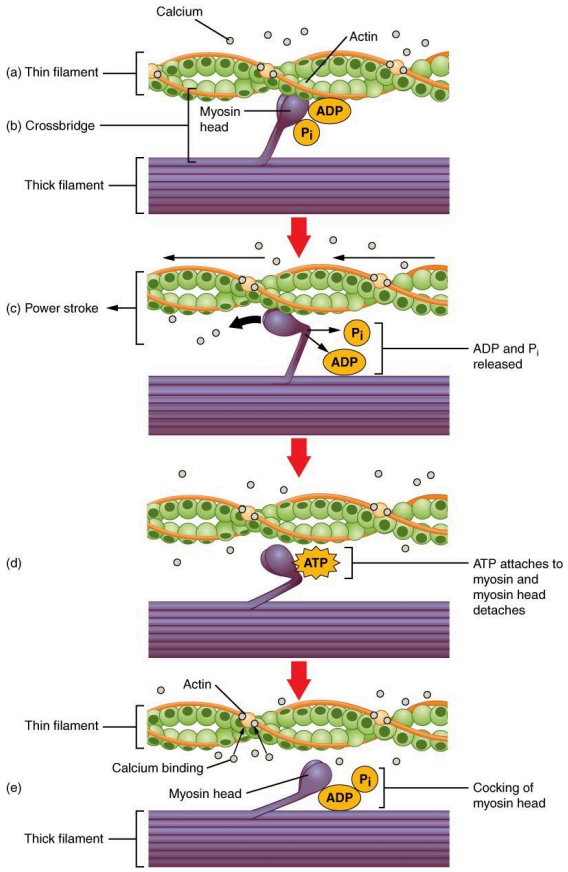
and half are pointing the other. Since the myosin heads on the opposite ends of the thick filaments all pull towards the middle, the overall effect is to cause the sarcomere to shorten. As all of the sarcomeres in the muscle fiber shorten, the entire muscle shortens or contracts.

9. In order for significant shortening of a skeletal muscle fiber to occur, the myosin heads must detach from the G-actin active sites and then re-attach to a different active site further along the neighboring actin molecule. This is rather similar to the fact that in order to climb a ladder, we must pull ourselves up a rung and then let go and move our hands and feet to higher rungs. In order for this release to occur, each myosin head must bind an ATP molecule. The binding of ATP to the myosin head allows it to release from the actin. The ATPase then hydrolyzes the ATP into ADP and a phosphate group, which causes the head to "re–cock" (the **recovery stroke),** preparing it for the next power stroke. Hence, binding ATP allows the head to release, and hydrolysis of ATP re–energizes the head for the next power stroke. During a single muscle cell contraction, each myosin molecule undergoes the entire cross-bridge cycle many times—a process known as **cross-bridge cycling**. As long as Ca2+ is present and the active sites are exposed, the process will continue.

One other important concept: Using the analogy above, when we climb a ladder, we don't take both hands off of the rungs at the same time. Likewise, when muscles contract, the myosin heads are cycling asynchronously, meaning that they don't all bind actin at the same time, and they don't all release at the same time. At any given time, the 300 or so myosin heads in one thick filament will be at different stages of the cross-bridge cycle.

The movement of myosin heads occurs in two phases:

1. The *power stroke* occurs when the myosin heads bend and ratchet the actin molecules past the myosin.
2. The *recovery stroke* involves the myosin heads detaching from actin and being cocked back into the high energy position to prepare for the next power stroke.

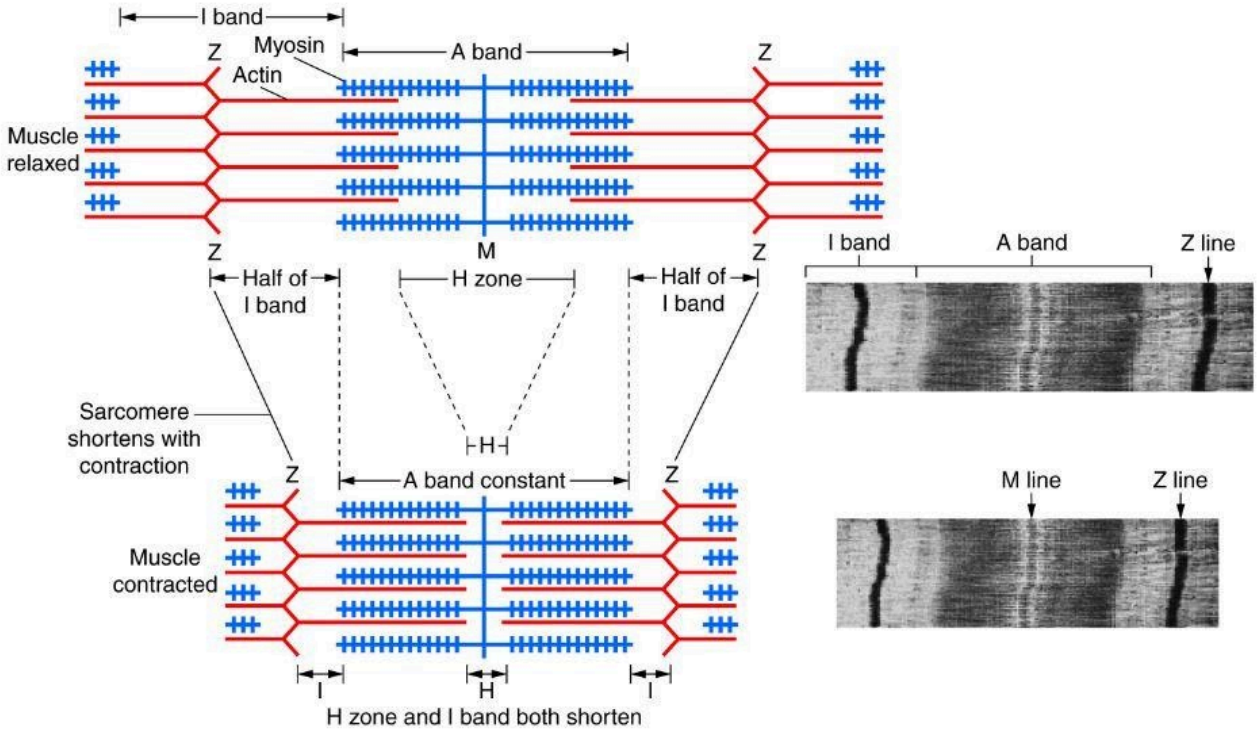


Skeletal Muscle Contraction. *Downloaded from Wikimedia Commons Dec 2013; Author: OpenStax College; Source:* https://books.byui.edu/-TZVB

(a) Calcium binds to Troponin and active site on actin exposed. (b) Myosin binds to Actin forming cross-bridge. (c) Phosphate released in a power stroke causing the myosin head to pivot and releasing ADP/Phosphate group released. (d) ATP attached to myosin head detaching cross bridge. (e) Myosin head hydrolyzed ATP to ADP and phosphate turning myosin back to ready position.

Relaxation begins when the release of acetylcholine ceases at the neuromuscular junction. The acetylcholine already in the synapse is broken down by acetylcholinesterase, ending action potential generation and propagation. This stops the release of Ca2+ from the sarcoplasmic reticulum (SR). Ca2+ ions diffuse away from troponin as the Ca2+ is actively transported back into the SR. This allows the troponin-tropomyosin complex to resume its resting position, blocking the active binding sites on the individual G-actin molecules. This prevents cross bridges from reforming and results in muscle relaxation. Even though each step begins the events of relaxation, the muscle will not fully relax until all calcium is pumped back into the SR.

To quickly review, the sliding filament model of muscle contraction explains the fact that when skeletal muscle fibers contract, the individual proteins (actin and myosin) don't shorten. Rather, they slide over each other. ATP is necessary for the detachment of myosin heads from actin. Notice also that when a sarcomere contracts, both the H zone and the light I band shrink in width, while the dark A band doesn't appear to narrow.



Sarcomere Anatomy. Author: courses.candelalearning.com;
Site: https://courses.candelalearning.com/olianp/chapter/muscular-levels-of-organization

7.3.2

Muscle Contractures and Cramps

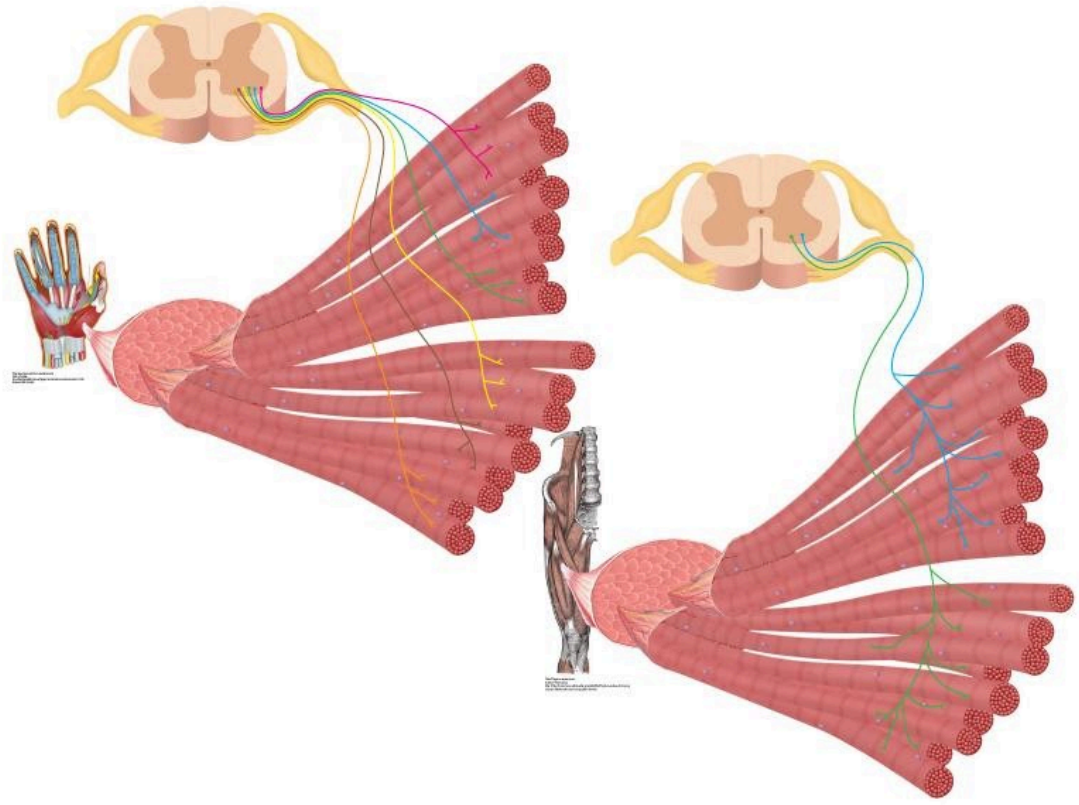
Muscle contracture is a term that generally refers to a muscle that has shortened and resists relaxing to its normal resting length. There are many possible causes of a contracture. Some of the reasons that tend to cause long term or even permanent contractures include prolonged immobilization, spasticity (spasm), and muscle weakness. Although rare, temporary contractures can occur with severe muscle fatigue, resulting in a condition called **physiologic contracture**.

Though not considered a "contracture" in the normal sense of the word, **rigor mortis** is muscle rigidity caused by a depletion of ATP. Rigor mortis is associated with death. It may take hours to fully develop, but cellular death results in the breakdown of the intracellular sarcoplasmic reticulum and the leakage of Ca^{2+}. The Ca^{2+} then initiates the events that allow cross bridge formation. However, cell death also results in the cessation of ATP production, and without ATP to cause the dissociation of myosin heads from actin, the muscle stays in a contracted position and cannot relax or be stretched.

Students often ask us what causes muscle cramps. It appears that there are many things that contribute, and physiologists are still puzzling this out. Most people have experienced or will experience a sudden, involuntary, and painful contraction at some point. It is very common during or after intense exercise. A characteristic that distinguishes cramps from contractures is that contractures are "electrically silent." This means that we do not see repeated action potentials coming down the motor neurons to the muscle cell. Contractures originate because of a physiological change of the muscle fiber itself and not the motor neuron that innervates it. Cramps, on the other hand, are associated with the repeated firing of action potentials in the motor neurons. Cramps are found most commonly in muscles of the leg, especially the lower leg and foot. The pathophysiology of cramps is poorly understood.

7.4

WHOLE MUSCLE CONTRACTION



Motor Neurons. *Image drawn by BYU-Idaho student Nate Shoemaker Spring 2017*

Skeletal Muscle Fiber Types

A Little Muscle Pharmacology

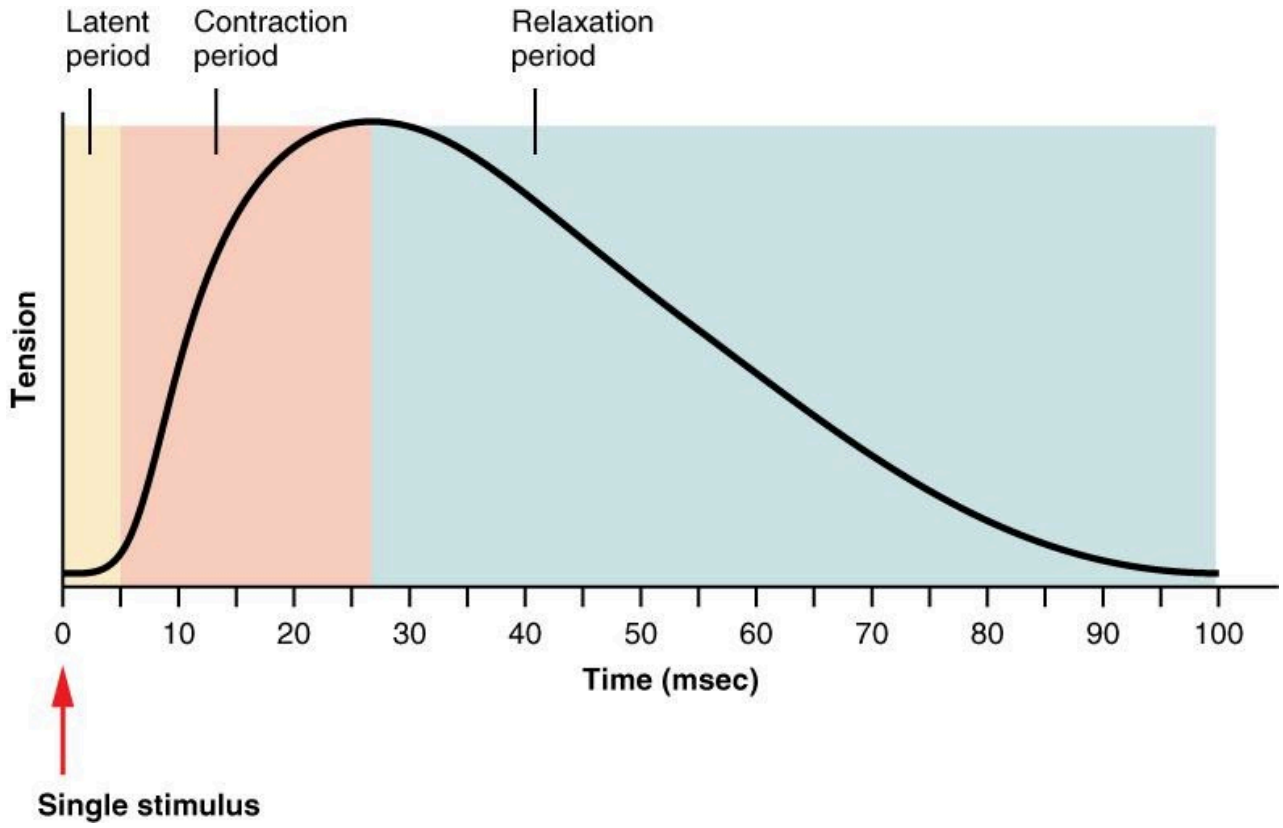
7.4.1

Motor Units

The motor neurons that innervate skeletal muscle fibers are called *alpha motor neurons*. As the alpha motor neuron enters a muscle, it divides into several branches, each innervating a muscle fiber (note this in the image above). One alpha motor neuron, along with all of the muscle fibers it innervates, is a *motor unit*. The size of the motor unit correlates with the function of the muscle. In muscles involved with fine, coordinated control, the motor units are very small with 3–5 muscle fibers per motor neuron. Muscles that control eye movement and muscles in our hands have relatively small motor units. On the other hand, in muscles involved with more powerful but less coordinated actions, like the muscles of the legs and back, the motor units are large with thousands of muscle fibers per motor neuron.

7.4.2

Physiology of a Muscle Twitch



Muscle Twitch Myogram. *Title: File:1012 Muscle Twitch Myogram.jpg; Author: OpenStax College; Site:https://commons.wikimedia.org/wiki/File:1012_Muscle_Twitch_Myogram.jpg; License: This file is licensed under the Creative Commons Attribution 3.0 Unported license.*

When an action potential travels down the motor neuron, it will result in a contraction of all of the muscle fibers associated with that motor neuron. The contraction generated by a single action potential is called a **muscle twitch**. A single muscle twitch has three components: the **latent period** or lag phase, the **contraction phase**, and the **relaxation phase**. The latent period is a short delay (1–2 msec) from the time when the action potential reaches the muscle until tension can be observed in the muscle. This is the time required for calcium to diffuse out of the SR and bind to troponin, the movement of tropomyosin off of the active sites, the formation of cross bridges, and the taking up of any slack that may be in the muscle. The contraction phase is when the muscle is generating tension and is associated with the cycling of cross bridges, and the relaxation phase is the time when the muscle returns to its normal length. The length of the twitch varies between different muscle types and could be as short as 10 msec (milliseconds) or as long as 100 msec (more on this later).

If a muscle twitch is just a single quick contraction followed immediately by relaxation, how do we explain the smooth continued movement of our muscles when they contract and move bones through a large range of motion? The answer lies in the ordering of the firing of the motor units. If all of the motor units fired simultaneously the entire muscle would quickly contract and relax, producing a very jerky movement. Instead, when a muscle contracts, motor units fire asynchronously, that is, one contracts and then a fraction of a second later another contracts before the first has time to relax and then another fires and so on. So, instead of a quick, jerky movement, the whole muscle contraction is very smooth and controlled. Even when a muscle is at rest, there is random firing of motor units. This random firing is responsible for what is known as **muscle tone**. So, a muscle is never "completely" relaxed, even when asleep. However, if the neuron to a muscle is cut, there will be no "muscle tone" and this is called flaccid paralysis. There are several benefits of muscle tone: First, it takes up the "slack" in the muscle so that when it is asked to contract, it can immediately begin to generate tension and move the limb. If you have ever towed a car you know what happens if you don't take the slack out of the tow rope before starting to pull. The second thing muscle tone does is deter muscle **atrophy**.

7.4.3

Types of Muscle Contraction

7.4.3 - Types of Muscle Contraction

Muscle contractions are described based on two variables: force (tension) and length (shortening). When the tension in a muscle increases without a corresponding change in length, the contraction is called an **isometric contraction** (iso = same, metric=length). Isometric contractions are important in maintaining posture or stabilizing a joint. On the other hand, if the muscle length changes while muscle tension remains relatively constant, then the contraction is called an **isotonic contraction** (tonic = tension). Furthermore, isotonic contractions can be classified based on how the length changes. If the muscle generates tension and the entire muscle shortens than it is a **concentric contraction**. An example would be curling a weight from your waist to your shoulder; the bicep muscle used for this motion would undergo a concentric contraction. In contrast, when lowering the weight from the shoulder to the waist the bicep would also be generating force but the muscle would be lengthening, this is an **eccentric contraction**. Eccentric contractions work to decelerate the movement at the joint. Additionally, eccentric contractions can generate more force than concentric contractions. Think about the large box you take down from the top shelf of your closet. You can lower it under total control using eccentric contractions but when you try to return it to the shelf using concentric contractions you cannot generate enough force to lift it back up. Strength training, involving both concentric and eccentric contractions, appears to increase muscle strength more than just concentric contractions alone. However, eccentric contractions cause more damage (tearing) to the muscle resulting in greater muscle soreness. If you have ever run downhill in a long race and then experienced the soreness in your quadriceps muscles the next day, you know what we are talking about.

Muscle size is determined by the number and size of the myofibrils, which in turn is determined by the amount of myofilament proteins. Thus, resistance training will induce a cascade of events that result in the production of more proteins. Often this is initiated by small, micro-tears in and around the muscle fibers. If the tearing occurs at the myofibril level the muscle will respond by increasing the amount of proteins, thus strengthening and enlarging the muscle, a phenomenon called hypertrophy. This tearing is thought to account for the muscle soreness we experience after a workout. As mentioned above, the repair of these small tears results in enlargement of the muscle fibers but it also results in an increase in the amount of connective tissue in the muscle. When a person "bulks up" from weight training, a significant percentage of the increase in the size of the muscle is due to increases in the amount of connective tissue. It should be pointed out that endurance training does not result in a significant increase in muscle size but increases its ability to produce ATP aerobically.

7.4.4

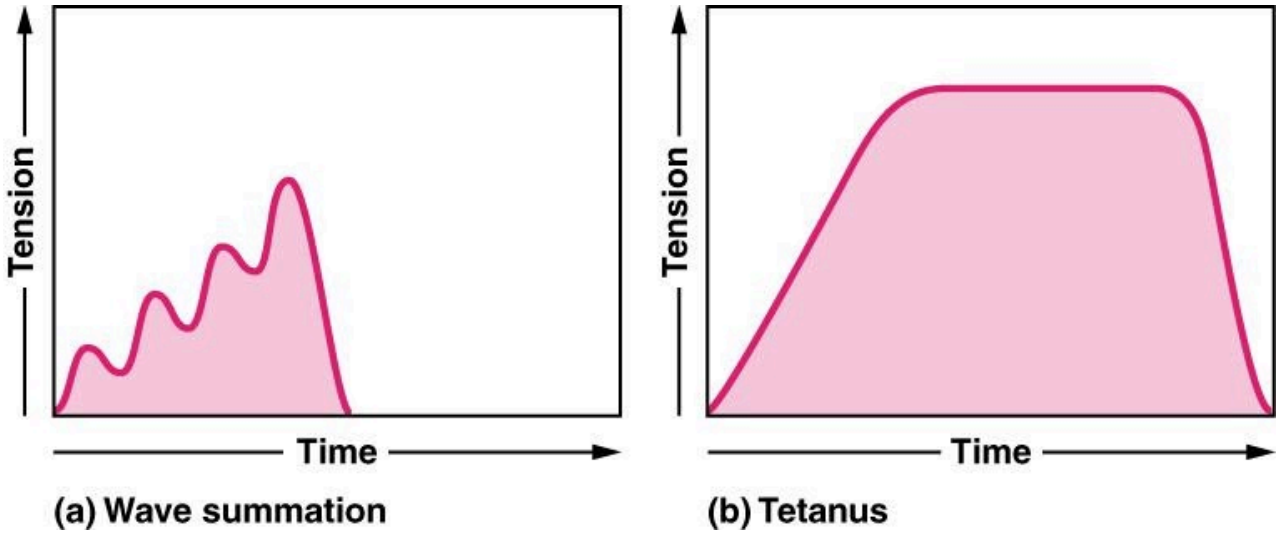
Factors That Influence the Force of Muscle Contraction

Obviously, our muscles are capable of generating differing levels of force during whole muscle contraction. Some actions require much more force generation than others; think of picking up a pencil compared to picking up a bucket of water. The question becomes, how can different levels of force be generated?

Multiple-Motor Unit Summation or Recruitment

It was mentioned earlier that all of the motor units in a muscle usually don't fire at the same time. One way to increase the amount of force generated is to increase the number of motor units that are firing at a given time. We say that more motor units are being **recruited.** The greater the load we are trying to move the more motor units that are activated. However, even when generating the maximum force possible, we are only able to use about 1/3 of our total motor units at one time. Normally they will fire asynchronously in an effort to generate maximum force and prevent the muscles from becoming fatigued. As fibers begin to fatigue they are replaced by others in order to maintain the force. There are times, however, when under extreme circumstances we are able to recruit even more motor units. You have heard stories of mothers lifting cars off of their children, this may not be total fiction. Watch the following clip to see how amazing the human body can be. Muscle recruitment. (Video Transcription Available)

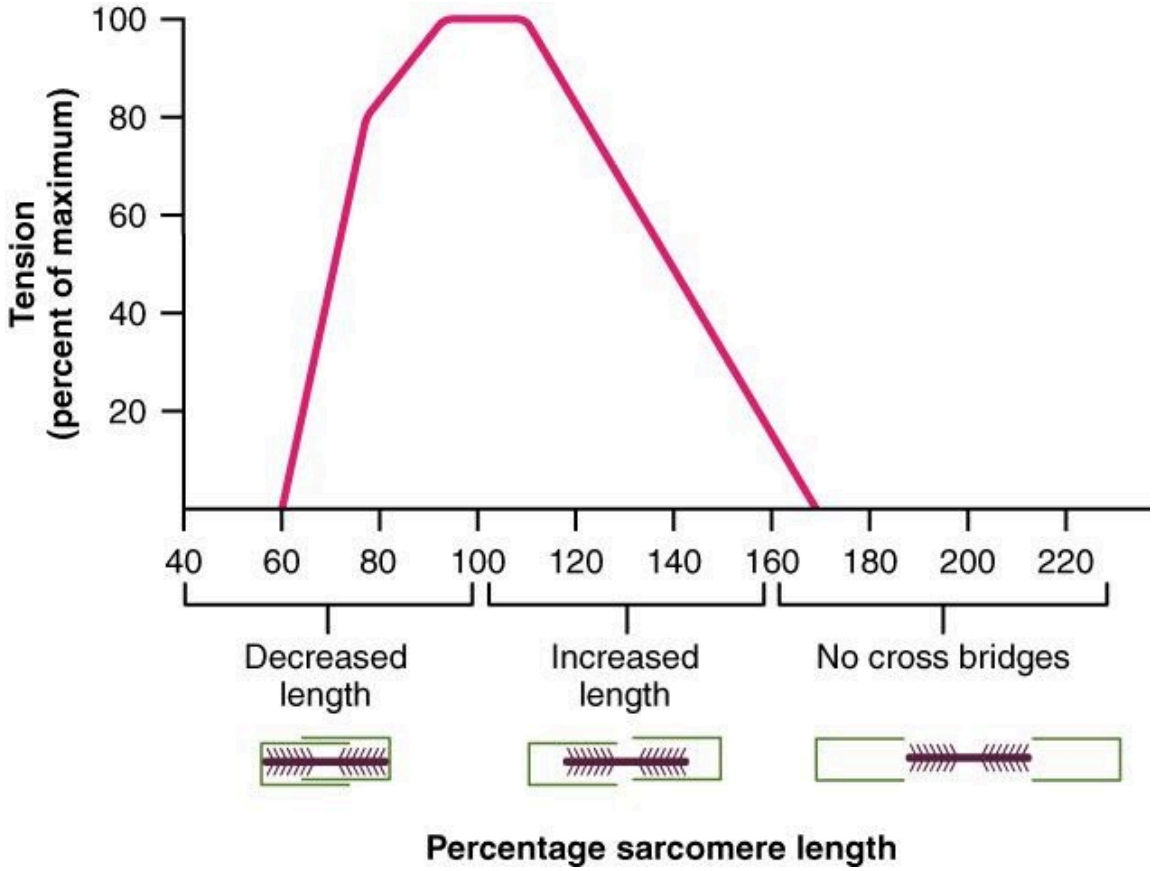
Wave Summation



Title: 1013_Summation_Tetanus.jpg; Author: OpenStax;
Site: http://cnx.org/contents/14fb4ad7-39a1-4eee-ab6e-3ef2482e3e22@6.6:67/Anatomy-&-Physiology;
License: This work is licensed by Rice University under a Creative Commons Attribution License License (3.0).

Recall that a muscle twitch can last up to 100 ms and that an action potential lasts only 1-2 ms. Also, with the muscle twitch, there is no refractory period so it can be re-stimulated at any time. If you were to stimulate a single motor unit with progressively higher frequencies of action potentials you would observe a gradual increase in the force generated by that muscle. This phenomenon is called **wave summation**. Eventually, the frequency of action potentials would be so high that there would be no time for the muscle to relax between the successive stimuli and it would remain totally contracted, a condition called **tetanus**. Essentially, with the high frequency of action potentials there isn't time to remove calcium from the cytosol. Maximal force, then, is generated with maximum recruitment, and; an action potential frequency sufficient to result in tetanus.

Length Tension Relationship



Muscle Length and Tension Graph. *Title: 1011_Muscle_Length_and_Tension.jpg; Author: OpenStax; Site: https://books.byui.edu/-LhYT;*
License: This work is licensed by Rice University under a Creative Commons Attribution License License (3.0).

It has been demonstrated experimentally that the starting length of the sarcomere influences the amount of force the muscle can generate. This observation has to do with the overlap of the thick and thin filaments. If the starting sarcomere length is very short, the thick filaments will already be pushing up against the Z-disc and there is no possibility for further sarcomere shortening, and the muscle will be unable to generate as much force. On the other hand, if the muscle is stretched to the point where myosin heads can no longer contact the actin, then again, less force will be generated. Maximum force is generated when the muscle is at a length that allows every myosin head to contact the actin and the sarcomere has the maximum distance to shorten. In other words, the thick filaments are at the very ends of the thin filaments. These data were generated experimentally using frog muscles that were dissected out and stretched between two rods. Intact muscles in our bodies are not normally stretched very far beyond their optimal length due to the arrangement of muscle attachments and joints.

However, you can do a little experiment that will help you see how force is lost when a muscle is in a very short or a very stretched position. This experiment will use the muscles that help you pinch the pad of your thumb to the pads of your fingers. These muscles are near maximal stretch when you extend your arm and also extend your wrist. As your wrist is cocked back into maximal extension, try to pinch your thumb to your fingers. See how weak it feels? Now, gradually flex your wrist back to a straight or neutral position. You should feel your pinch get stronger. Now, flex your elbow and your wrist. With your wrist in maximal flexion, the muscles you use to pinch with are near their most shortened position. Try pinching again. It should feel weak. But, again, as you extend your wrist back to neutral you should feel your pinch get stronger.

7.4.5

Energy Source for Muscle Contraction

7.4.5 - Energy Source for Muscle Contraction

The ultimate source of energy for muscle contraction is ATP. Recall that each cycle of a myosin head requires an ATP molecule. Multiply that by all of the myosin heads in a muscle and the number of cycles each head completes each twitch and you can start to see how much ATP is needed for muscle function. It is estimated that we burn approximately our entire body weight in ATP each day so it becomes apparent that we need to constantly replenish this important energy source. For muscle contraction, there are four ways that our muscles get the ATP required for contraction.

1. **Cytosolic ATP:** This ATP represents the "floating" pool of ATP, or that which is present and available in the cytoplasm. This ATP requires no oxygen (anaerobic) to make it (because it is already there) and is immediately available but it is short lived. It provides enough energy for a few seconds of maximal activity in the muscle-not the best source for long term contraction. Nevertheless, for the muscles of the eyes that are constantly contracting quickly but for short periods of time, this is a great source.

2. **Creatine Phosphate:** Once the cytosolic stores of ATP are depleted, the cell calls upon another rapid energy source, Creatine Phosphate. Creatine phosphate is a high energy compound that can rapidly transfer its phosphate to a molecule of ADP to quickly replenish ATP without the use of oxygen. This transfer requires the enzyme creatine kinase, an enzyme that is located on the M-line of the sarcomere. Creatine phosphate can replenish the ATP pool several times, enough to extend muscle contraction up to about 10 seconds. Creatine Phosphate is the most widely used supplement by weight lifters. Although some benefits have been demonstrated, most are very small and limited to highly selective activities.

3. **Glycolysis:** Glycolysis, as the name implies, is the breakdown of glucose. The primary source of glucose for this process is from glycogen that is stored in the muscle. Glycolysis can function in the absence of oxygen and as such, is the major source of ATP production during anaerobic activity. This series of chemical reactions will be a major focus in the next unit. Although glycolysis is very quick and can supply energy for intense muscular activity, it can only be sustained for about a minute before the muscles begin to fatigue.

4. **Aerobic or Oxidative Respiration:** The mechanisms listed above can supply ATP for maybe a little over a minute before fatigue sets in. Obviously, we engage in muscle activity that lasts much longer than a minute (things like walking or jogging or riding a bicycle). These activities require a constant supply of ATP. When continuous supplies of ATP are required, the cells employ metabolic mechanisms housed in the mitochondria that utilize oxygen. We normally refer to these processes as aerobic metabolism or oxidative metabolism. Using these aerobic processes, the mitochondria can supply sufficient ATP to power the muscle cells for hours. The down side of aerobic metabolism is that it is slower than anaerobic mechanisms and is not fast enough for intense activity. However, for moderate levels of activity, it works great. Although glucose can also be utilized in aerobic metabolism, the nutrient of choice is fatty acids. As described below, slow-twitch and fast-twitch oxidative fibers are capable of utilizing aerobic metabolism.

7.4.6

Fatigue

When we think of skeletal muscles getting tired, we often use the word fatigue, however, the physiological causes of fatigue vary considerably. At the simplest level, fatigue is used to describe a condition in which the muscle is no longer able to contract optimally. To make discussion easier, we will divide fatigue into two broad categories: **Central fatigue** and **peripheral fatigue**. Central fatigue describes the uncomfortable feelings that come from being tired, it is often called "psychological fatigue." It has been suggested that central fatigue arises from factors released by the muscle during exercise that signal the brain to "feel" tired. Psychological fatigue precedes peripheral fatigue and occurs well before the muscle fiber can no longer contract. One of the outcomes of training is to learn how to overcome psychological fatigue. As we train we learn that those feelings are not so bad and that we can continue to perform even when it feels uncomfortable. For this reason, elite athletes hire trainers that push them and force them to move past the psychological fatigue.

Peripheral fatigue can occur anywhere between the neuromuscular junction and the contractile elements of the muscle. It can be divided into two subcategories, **low frequency** (marathon running) and **high frequency** (circuit training) fatigue. High-frequency fatigue results from impaired membrane excitability as a result of imbalances of ions. Potential causes are inadequate functioning of the Na^+/K^+ pump, subsequent inactivation of Na^+ channels and impairment of Ca^{2+} channels. Muscles can recover quickly, usually within 30 minutes or less, following high-frequency fatigue. Low-frequency fatigue is correlated with impaired Ca^{2+} release, probably due to excitation coupling contraction problems. It is much more difficult to recover from low-frequency fatigue, taking from 24 hours to 72 hours.

In addition, there are many other potential fatigue contributors, these include accumulation of inorganic phosphates, hydrogen ion accumulation and subsequent pH change, glycogen depletion, and imbalances in K^+. Please note that a factor not on the list is lactic acid as it does not contribute to fatigue or muscle soreness. The reality is we still don't know exactly what causes fatigue and much research is currently devoted to this topic.

7.4.7

Skeletal Muscle Fiber Types

Classically, skeletal muscle fibers can be categorized according to their speed of contraction and their resistance to fatigue. These classifications are in the process of being revised, but the basic types include:

1. *Slow twitch oxidative (type I) muscle fibers,*
2. *Fast-twitch oxidative-glycolytic (Type IIA) muscle fibers, and*
3. *Fast-twitch glycolytic (Type IIX) fibers.*

Fast-twitch (type II) fibers develop tension two to three times faster than slow-twitch (type I) fibers. How fast a fiber can contract is related to how long it takes for completion of the cross-bridge cycle. This variability is due to different varieties of myosin molecules and how quickly they can hydrolyze ATP. Recall that it is the myosin head that splits ATP. Fast-twitch fibers have a more rapid ATPase (splitting of ATP into ADP + Pi) ability. Fast-twitch fibers also pump Ca^{2+} ions back into the sarcoplasmic reticulum very quickly, so these cells have much faster twitches than the slower variety. Thus, fast-twitch fibers can complete multiple contractions much more rapidly than slow-twitch fibers. For a complete list of how muscle fibers differ in their ability to resist fatigue see the table below:

	Slow Twitch Oxidative (Type I)	Fast-twitch Oxidative (Type IIA)	Fast-Twitch Glycolytic (Type IIX)
Myosin ATPase activity	Slow	fast	fast
Size (diameter)	Small	medium	large
Duration of contraction	Long	short	short
SERCA pump activity	Slow	fast	fast
Fatigue	Resistant	resistant	easily fatigued
Energy utilization	aerobic/oxidative	both	anaerobic/glycolytic
capillary density	High	medium	low
mitochondria	high numbers	medium numbers	low numbers

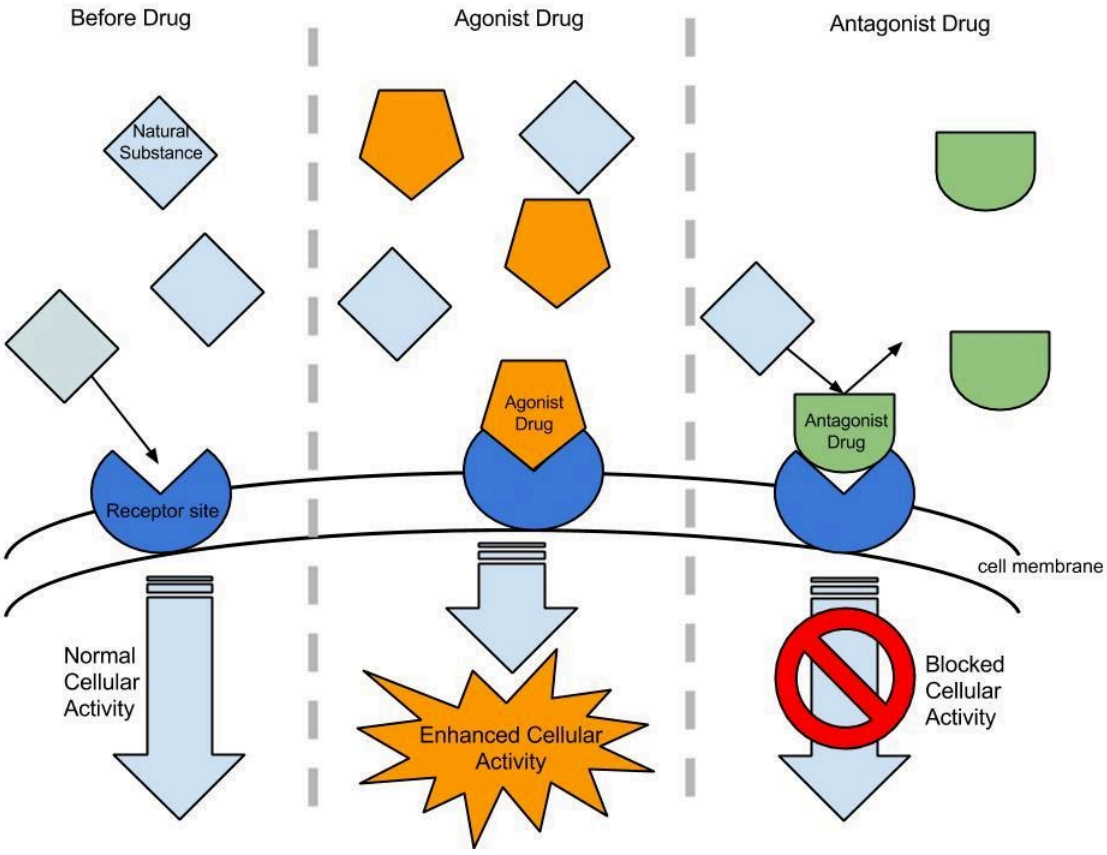
Color	red (contain myoglobin)	red (contain myoglobin)	white (no myoglobin)

In human skeletal muscles, the ratio of the various fiber types differs from muscle to muscle. For example, the gastrocnemius muscle of the calf contains about half slow and half fast type fibers, while the deeper calf muscle, the soleus, is predominantly slow twitch. On the other hand, the eye muscles are predominantly fast twitch. As a result, the gastrocnemius muscle is used in sprinting while the soleus muscle is important for standing. In addition, women seem to have a higher ratio of slow twitch to fast twitch compared to men. The "preferred" fiber type for sprinting athletes is the fast-twitch glycolytic, which is very fast, however, most humans have a very low percentage of these fibers, < 1%. Muscle biopsies of one world class sprinter revealed 72% fast twitch fibers and amazingly 20% were type IIX. The Holy Grail of muscle research is to determine how to change skeletal muscle fibers from one type to another. It appears that muscle fiber types are determined embryologically by the type of neuron that innervates the muscle fiber. The default muscle appears to be slow, type I fibers. If a muscle is innervated by a small neuron that muscle fiber will remain slow, whereas large myelinated fibers induce the fast isoforms. In addition, the frequency of firing rates of the neuron also alters the muscle fiber type. Research suggests that humans have subtypes of fibers, making up about <5% of the muscle, that are dually innervated and allow for switching between slow and fast to occur. Generally, it would appear that genetics determine the type of innervation that occurs and subsequent muscle fiber types and that training may be able to slightly alter the ratios due to the dually innervated muscles. However, since <5% have dual innervation, genetics is going to play a much greater role in your fiber types than your training.

7.4.8

A Little Muscle Pharmacology

At this point, you may be thinking, “how on earth did anyone figure out this stuff?” To answer this question you need to understand that scientists who study physiological processes often employ drugs. Yep, drugs! A drug that has the same effect as acetylcholine will result in all kinds of information about how something works. We use the terms agonist and antagonist when referring to drugs. A drug that has the same effect as the neurotransmitter (acetylcholine) would be considered an agonist. A drug that blocks the effect of the neurotransmitter is called an antagonist.



Agonist & Antagonist. Author: Dolleyj Site: https://books.byui.edu/-WEWT;
License: CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons.

Look at the table below of different drugs used in the study of muscle physiology and see if you can predict the effect the drug would have on a muscle.

Class of Drug	Example	Method of Action	Result on Muscle
Neuromuscular blocker	tubocurarine (chemical obtained from the bark of a South American plant, used as arrow poison); alpha bungarotoxin (snake poison), pancuronium (lethal injection drug)	Acetylcholine receptor antagonist	Flaccid paralysis
Neuromuscular blocker	Succinylcholine (a synthesized chemical, known as the “perfect poison” for murder)	Acetylcholine receptor antagonist (initial depolarization but then blocks the receptor)	Flaccid paralysis
Neuromuscular junction	Neostigmine (a synthesized chemical)	Inhibits Acetylcholinesterase activity	Spastic paralysis
Contractility	Salbutamol (a synthesized chemical also known as albuterol)	Enhances SERCA pump activity	Reduced contractility
Contractility	Caffeine (chemical found in seeds, nuts or leaves, used as an insecticide by the plants)	Enhances Ca++ release at the sarcoplasmic reticulum	Increased contractility
Neuromuscular junction	Botulism	Blocks SNARE proteins	Flaccid Paralysis
Neuromuscular junction	Latrotoxin (Black widow spider poison)	SNARE protein agonist	Spastic paralysis

8.0

MODULE 8: METABOLISM

8.1

ENERGTY CYCLE, ATP and ELECTRON CARRIERS

"There is no easy way out. If there were, I would have bought it. And believe me, it would be one of my favorite things."

-Oprah Winfrey

Have you ever noticed that in our world there seems to be a never ending parade of "fads?" This is especially true with diet and exercise. Such "fads" promise quick weight loss, instant body shaping or both. Weeding out false information and gimmicks can be more difficult than it may seem. There are so many testimonials, "evidence," and promises that clever advertisers can leave you feeling that you are the only one not taking advantage of a "sure thing." Ultimately, it is just too tempting for many and soon we are shelling out hard earned dollars for something that is here today, gone tomorrow and we are little changed from the experience. Weight loss and fitness gimmicks make up a billion dollar industry which leaves us with an awful lot of pills, powders and plans to sort through.

A few years ago, I found myself fascinated by an article in Time magazine. (https://books.byui.edu/-CZeE) This article suggests that a type of fat exists that can make us thin. It is called "brown fat." This article and others like it claim that the activation of Brown Fat would cause weight loss with no need for diet or exercise. Evidence that this might work comes from research that shows mice getting skinny when brown fat is activated in their bodies. Also, it has been shown that many humans who maintain a lean physique have higher quantities of Brown Fat. Researchers tell us that as little as 50g of brown fat can consume as much as 20% of the total daily caloric expenditure (Lidell and Enerback, 2010).

You must have some questions by now. Wouldn't you like to know what brown fat is? How does it work? Does it work? Why don't we hear about medical treatments that use it? How does brown fat burn energy instead of store it? What makes fat brown instead of white?

You know that the media has an insatiable appetite to advertise the next "greatest thing" with regard to diet and exercise. Your ability to react intelligently to this type of information will be greatly enhanced through a more solid understanding of Metabolism.

Metabolism in its most simplified "big picture" description is largely a story of chemical bonds, more specifically the electrons in these bonds. You probably recall that electrons are a subatomic particle with a negative charge. Electrons also move with considerable energy and speed as they find themselves drawn to protons (which are positive and found in atomic nuclei). So, what if we could take some high energy electrons and store them somewhere. Later, when we wanted to harvest some of the electron energy to do some work we could go get them. This is kind of how a battery works. Smart people have figured out how to store high energy electrons and when we want these electrons to do something (like flow through a flashlight filament) we can get them and use them. Living things also use high energy electrons to power the processes of biology. We as humans don't plug ourselves into a Duracell battery, but we do eat foods that have stored high energy electrons.

Our story starts with the sun. The earth gets energy from the sun. This energy is called solar energy. You can feel some of this energy when you go out on a warm day. Solar energy is absorbed by plants and when a plant gets solar energy it

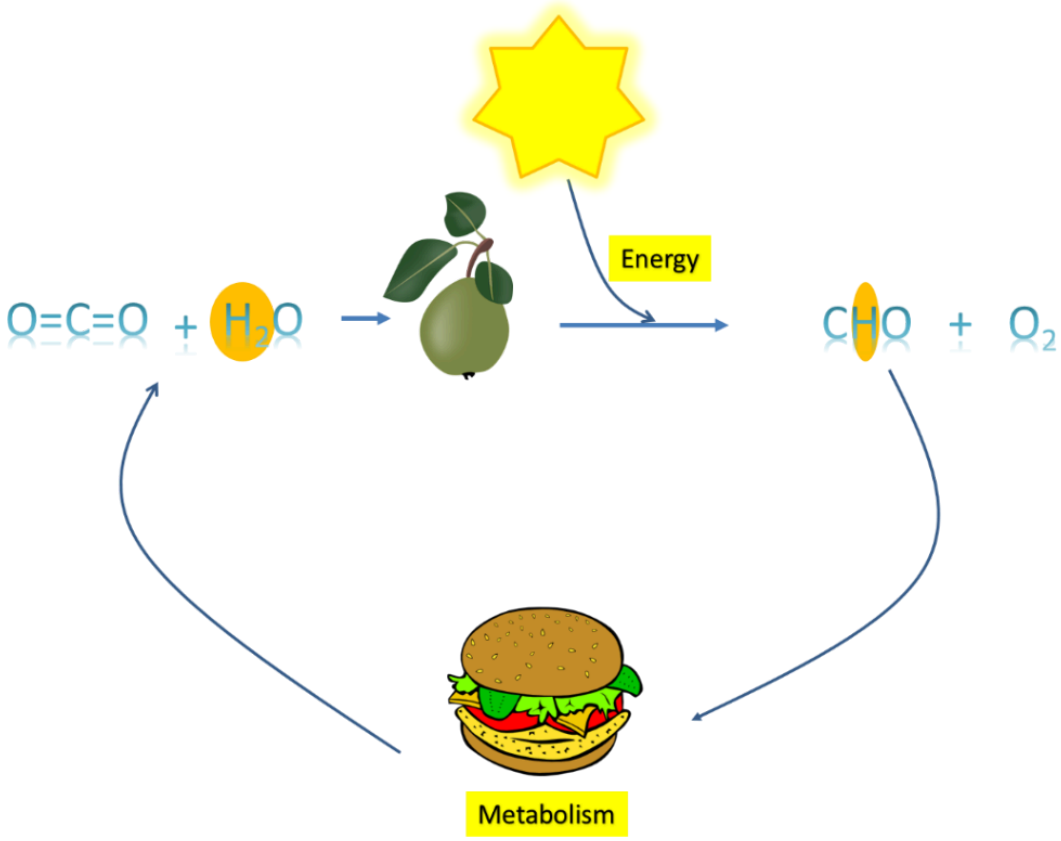
uses this energy to excite electrons.

An electron in a bond between oxygen and hydrogen has a certain amount of energy. An electron in a bond between carbon and hydrogen has even a greater amount of energy. When sunlight strikes a plant, the solar energy is used to break an oxygen-hydrogen bond and then create a carbon-hydrogen bond. In this very incredible and complex process called photosynthesis, plants use water as the source of oxygen-hydrogen bonds and carbon dioxide (CO_2) serves as the source of carbon that will accept a hydrogen with its newly energized electron. Recall that whenever an atom receives more electrons, we say that the atom has been **reduced**, and whenever an atom loses electrons, we say that the atom has been **oxidized**. In our story of photosynthesis, CO_2 is reduced and H_2O is oxidized.

Let's summarize to this point. Plants take CO_2 and H_2O, they use solar energy to excite some electrons in H_2O and then transfer the excited electron with a hydrogen to carbon. Now, we have a higher energy C-H bond. Reduced CO_2 can be transformed into a variety of molecules with different numbers of C-H bonds. These molecules make up the sugars, lipids and proteins that you have learned about. If you were to look up the chemical structures of any sugars, lipids, or proteins, you would see a lot of C-H bonds.

So, in a way, plants create a kind of "battery." The carbon-hydrogen bonds formed by plants can exist for a long time as sugars, lipids and proteins. If the high energy electrons are allowed to go back and form an O-H bond, energy would have to be released. This is what our body does. Our cells have the ability to facilitate a transfer of high energy electrons in C-H bonds back to O-H bonds and use the energy that is released.

Now, you hopefully see this cycle of energy. We eat C-H bonds when we consume carbohydrates, lipids and proteins. In our cells we "process" the C-H bonds in a way that allows us to put the high energy electrons back onto Oxygen. Energy is released and we use the energy to run the cells of our body. Also, if you note in figure 1, CO_2 and H_2O are created again when we allow high energy electrons to return back to oxygen. Plants use the CO_2, and the H_2O and the cycle continues.



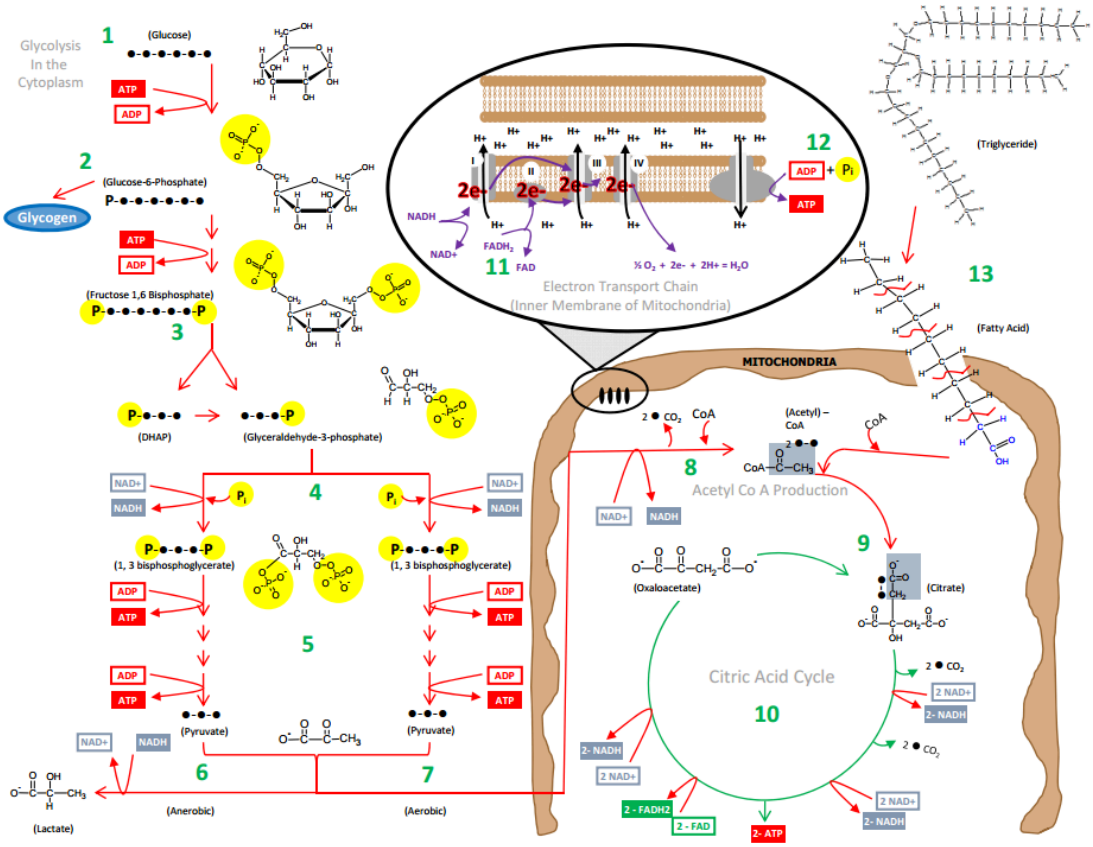
Energy Cycle. *Image created by JS at BYU Idaho. Clipart from clker.com; License Public Domain;http://www.clker.com/clipart-4079.html and http://www.clker.com/clipart-stew-pear-with-leafs.html*

The image above is a graphical representation of the energy cycle. Solar energy from the sun excites electrons enough that they can facilitate a bond with carbon instead of oxygen (photosynthesis). We eat these C-H bonds in carbohydrates, lipids and proteins. Our body cells allow the high energy electrons in C-H bonds to return to a lower energy state in O-H bonds and we produce CO_2 and H_2O, which can cycle back to be used in photosynthesis again.

Bioenergetics is the study of how energy is transferred through the chemical reactions of living systems. Our cells are always breaking and making bonds. Every time a bond is broken, energy is released. Every time a bond is made energy is required in forming that bond. Keep in mind that whenever we see a reaction that involves the synthesis of new molecules, we call this an **Anabolic** reaction (think "anabolic steroids" and it shouldn't surprise you that these steroids stimulate reactions that make new proteins in muscle cells). **Catabolic** reactions occur when molecules are broken down into smaller and smaller parts. Several catabolic reactions occur in our cells to break sugars, proteins, and lipids down into smaller and smaller parts until finally the C-H bonds have been processed to allow the energy in such bonds to be used in the form of ATP.

Metabolism refers to all of the catabolic and anabolic processes that a cell is engaged in. The details of metabolism can be daunting and normally require several classes in biology and chemistry to understand all that is known about metabolic processes. In this class, we hope to provide a simplified overview of metabolism and show you the "big picture" of how our cells capture energy inherent to electrons in carbon-hydrogen bonds and then use that energy to make ATP.

The main metabolic processes that create energy in our bodies that will be summarized in this module are Glycolysis (the breakdown of sugars), the Citric Acid Cycle (cellular respiration), the Electron Transport Chain (where ATP is created due to the transfer of electrons and protons along the membrane of the mitochondria during cellular respiration), Lipolysis and Beta Oxidation (the breakdown of fats), and Protein Metabolism. Below you will see an image that illustrates all of these metabolic processes as they work together to create energy. Refer to this image as you learn about the details of metabolism.



The “Big Picture” of Metabolism:
Glycolysis, Citric Acid (Krebs) Cycle, Electron Transport Chain, Beta Oxidation and Lipolysis.

Image created at BYU-Idaho by JS 2010

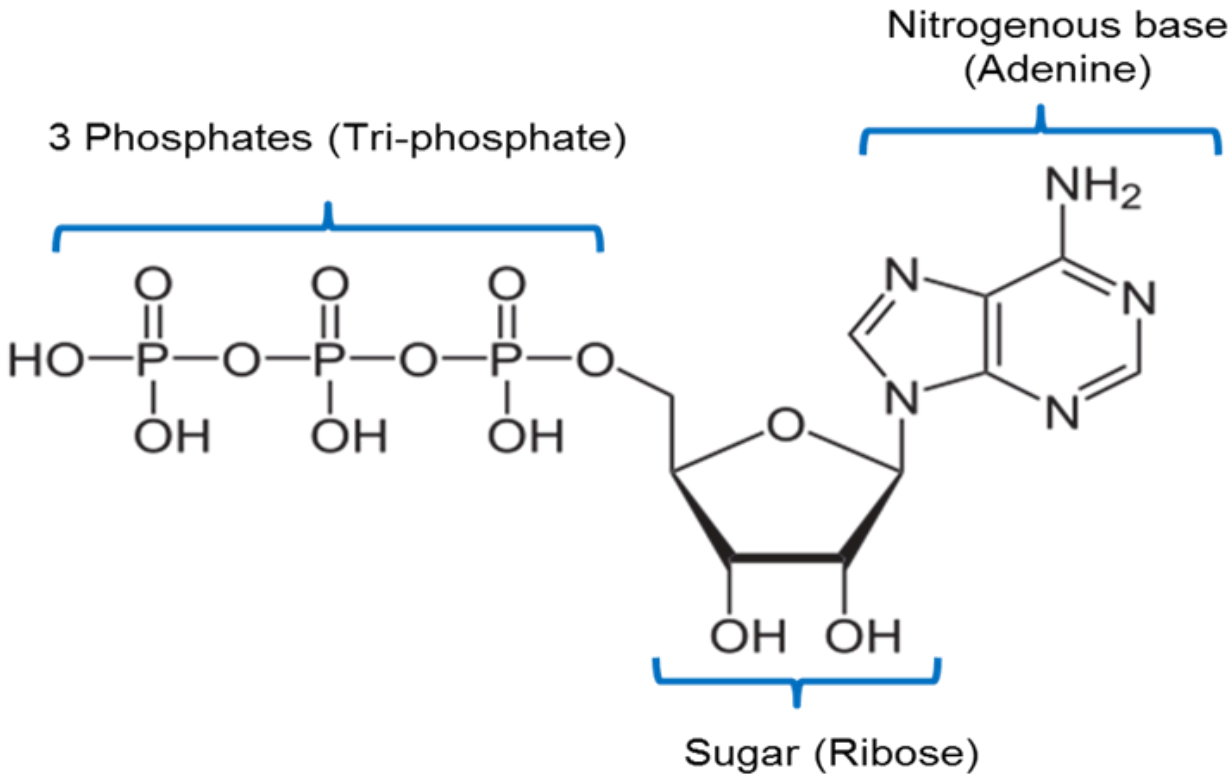
*** You can find a detailed description of each metabolic process shown in this image at the end of each section below.**
You can also download this image and the complete summary at this link: Metabolism Summary.

ATP
Electron Carriers (NAD and FAD)

8.1.1

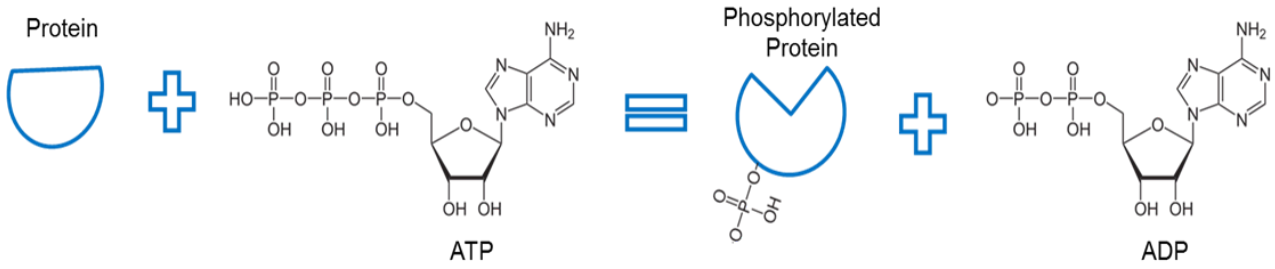
ATP

Cells use a molecule called **Adenosine Triphosphate** (or ATP) as an energy source. The phosphates in this molecule can supply energy to substrates in our cells. Enzymes exist in our cells that can remove a phosphate from ATP and attach it to a different molecule-usually a protein. When this happens, we say that the protein has been phosphorylated. Think of the third phosphate as being a little sack of energy. When it is transferred to a protein, this energy can be used to do something. For example, in the figure just below, the protein changes its shape when it becomes phosphorylated. When proteins change their shape, we often call this a conformational change to the protein structure. There are many proteins in the body that use a phosphate from ATP to induce a conformational change. This shifting of the protein shape ultimately allows for things like muscle contraction, cell mobility, membrane transport, and enzyme action. Cells and life exist only if a consistent and steady supply of ATP is available.



Chemical Structure of ATP. Image created by JS at BYU Idaho Fall 2013.

The image above is a representation of the chemical structure of ATP. ATP includes a nitrogenous base called adenine joined to a 5 carbon sugar called ribose and 3 phosphate groups.



Phosphorylation. Image created by JS at BYU-Idaho Fall 2013.

ATP is used to phosphorylate a protein. An enzyme, called a kinase (not shown) removes a phosphate from ATP and facilitates a bond between the phosphate and some other protein. The bonding of a phosphate to a protein in this manner is called phosphorylation. The phosphate bond with the protein has higher energy. Notice that phosphorylation uses this energy to cause a conformational change of the protein shape.

8.1.2

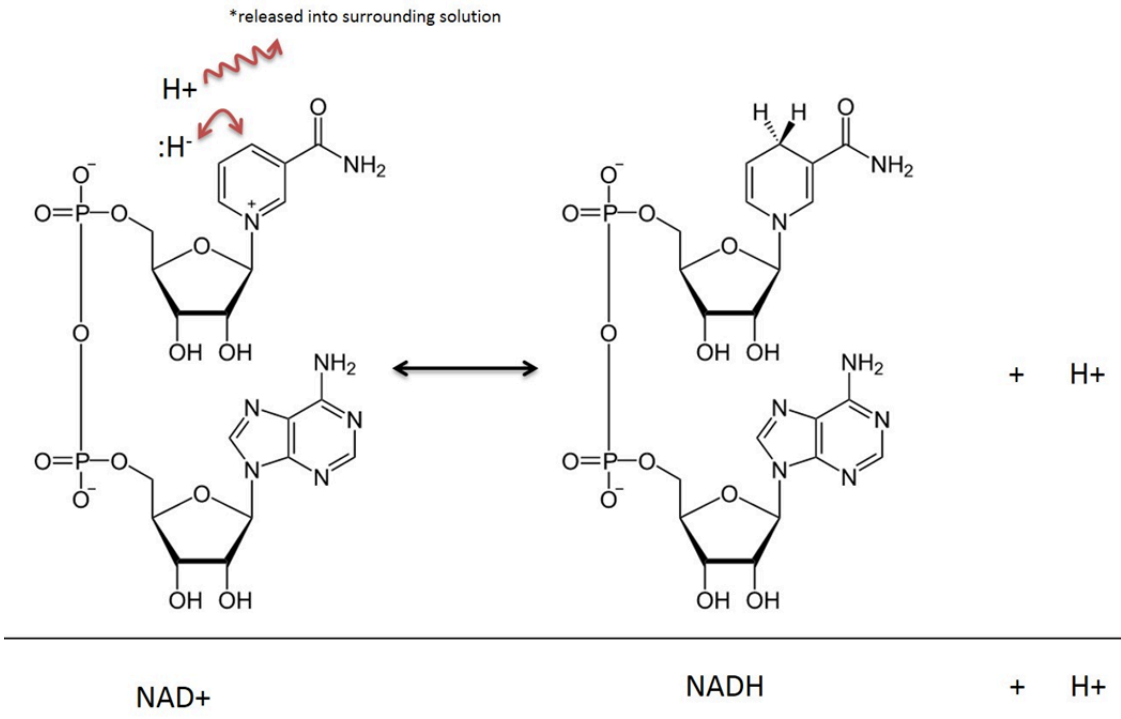
Electron Carriers (NAD and FAD)

Nicotinamide Adenine Dinucleotide (NAD) and **Flavin Adenine Dinucleotide** (FAD) are coenzymes involved in reversible oxidation and reduction reactions. It is often stated that these compounds are electron carriers because they accept electrons (become reduced) during catabolic steps in the breakdown of organic molecules such as carbohydrates and lipids. Then, these reduced coenzymes can donate these electrons to some other biochemical reaction normally involved in a process that is anabolic (like the synthesis of ATP).

NAD^+ / NADH

Nicotinamide Adenine Dinucleotide in its oxidized state is called **NAD^+**, after being reduced (or accepting electrons), it is referred to as **NADH**. The vitamin Niacin (also called B3) is used to derive this compound. Niacin provides the organic ring structure that will directly participate in the transfer of a hydrogen atom and 2 electrons. NAD^+ is often found in conjunction with a "*dehydrogenase*" enzyme. A dehydrogenase reaction removes two hydrogen atoms; one as a hydride (H^-) (*a hydride is a hydrogen atom with 2 electrons*) and one as a hydrogen cation (H^+) (*and of course, a hydrogen cation has no electrons*). The hydride bonds with NAD^+ and creates a reduced compound of Nicotinamide Adenine Dinucleotide (NADH). The second hydrogen atom (H^+) is released into solution.

As you examine the reactions for metabolism, look for reactions that yield NADH. NADH will be important as it will deliver the hydrogens and electrons that it picks up to biochemical processes that can use the electrons and hydrogens to make ATP.



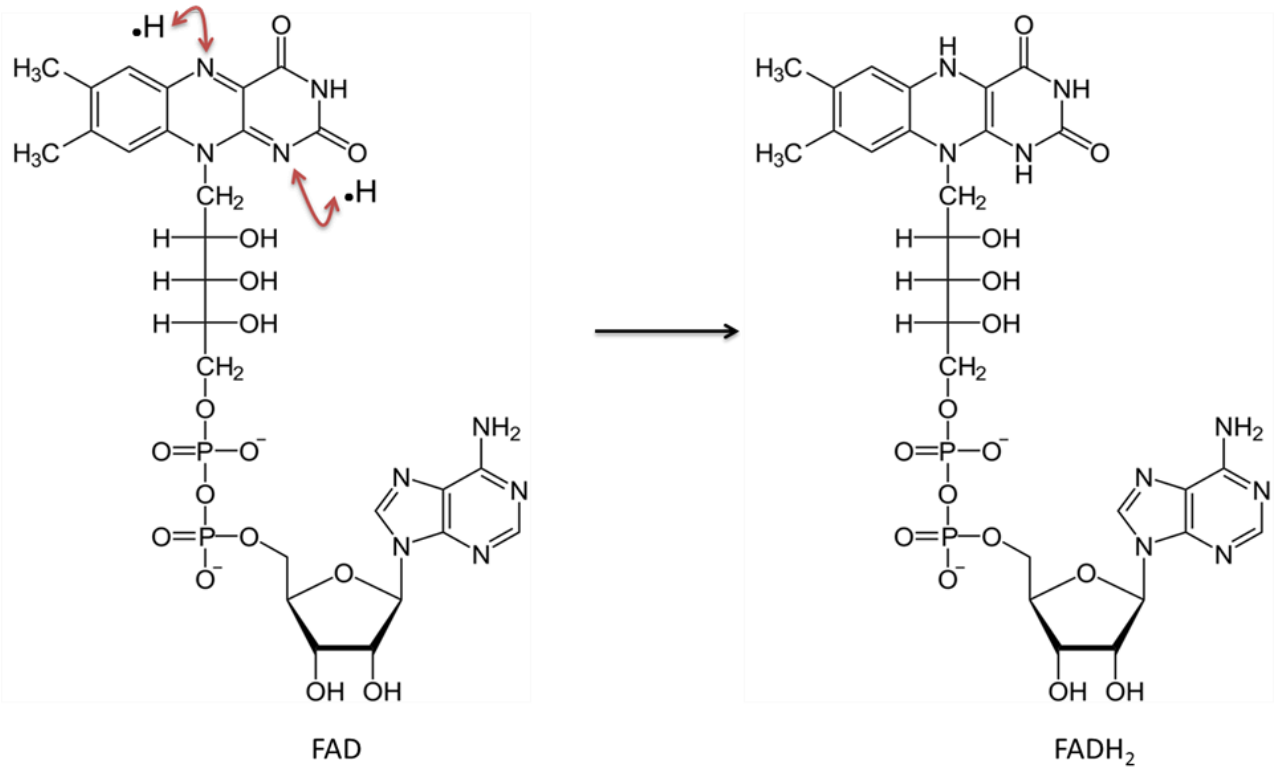
NAD+ Reduction to NADH + H^+. Image created by JS at BYU-Idaho Fall 2013.

In metabolic reactions that involve NAD, two hydrogen atoms and two electrons are removed from a substrate and transferred to NAD^+. NAD^+ accepts a hydride ion (a hydrogen with 2 electrons) and becomes Nicotinamide Adenine Dinucleotide in the reduced form (NADH). The hydrogen cation that is also captured in the reaction is released into the surrounding solution. Remember that this reaction is reversible.

In the explanation of reactions that occur in Metabolism, it is common to ignore the H^+ released into solution and this text will depict the outcome of NAD reduction as simply NADH, rather than NADH + H^+.

FAD / $FADH_2$

Flavin adenine dinucleotide in its oxidized state is called FAD. After being reduced, it is called $FADH_2$. See figure 5 in the metabolism summary image for a molecular illustration. The vitamin, riboflavin (or B2) is used to derive this compound. Riboflavin provides the ring structures that will directly participate in the transfer of two hydrogen atoms (each with one electron this time). Similar to NAD, FAD works in association with a "*dehydrogenase*" enzyme. The reaction removes two hydrogen atoms; each a proton with one electron. Both hydrogen atoms bond with FAD. This reaction does not release an H^+ into solution like the reduction of NAD does.



FAD Conversion to FADH2. Image created by JS at BYU Idaho F2013.
Flavin adenine dinucleotide in the oxidized form (FAD) accepts two hydrogen atoms (each with one electron) and becomes $FADH_2$.

As you examine the reactions for metabolism, look for a reaction that yields $FADH_2$. Similar to NADH, $FADH_2$ will be important as it will deliver hydrogens and electrons to biochemical processes that can use the electrons and hydrogens to make ATP.

8.2

GLYCOLYSIS

Glycolysis literally means the breakdown of sugar (Glyc = sugar or sweet and Lysis = to cut or loosen). Glycolysis occurs in the cytoplasm of the cell. Beginning on the next page, you will find depictions of the step by step biochemical reactions that make up Glycolysis. In short, glycolysis takes 1 glucose molecule of 6 carbons and makes two 3 carbon molecules called pyruvate. In the process, electrons and hydrogen atoms are captured by NAD^+. Any energy liberated will be released as heat, or captured as ATP or NADH.

Molecule	Net Yield through Glycolysis
ATP	2
NADH	2
Pyruvate	2

Glycogenesis

It should be noted that when there is a surplus of glucose, some of the glucose will be converted to glycogen. Later, when blood sugar begins to fall, glycogen can be broken down into monomers of glucose again, which will enter glycolysis. Glycogen synthesis is an important process to help us store" sugar for use when we are not eating but need to maintain appropriate blood sugar levels.

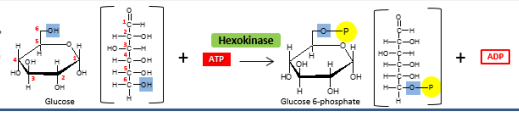
We have next a very detailed figure for the 10 steps of glycolysis. It has more information than you will need for the exam (remember, the exam will test you at the level of the "summary" figure that you got previously). However, this image is nice to have as a reference and can help you if you find videos or information in your online research that uses terminology not found in the summary figure.

10 STEPS OF GLYCOLYSIS

For each step, the part of the molecule that actually changes is highlighted in blue. Enzymes that catalyze the reaction are highlighted in green. Although we recognize that sugars exist in a ring form within the cell, sugars are shown in brackets in their open chain form to assist visualization of the numbered carbons. Finally, note that the first 5 steps are "energy requiring". Energy will not be harvested until the steps on the next page.

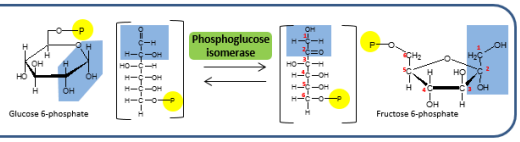
Step 1

Glucose enters the cell and ATP is used to phophorylate it. This fist step requires energy in the form of ATP. The polar nature of the phosphate group prevents passage of the Glucose back out of the cell.



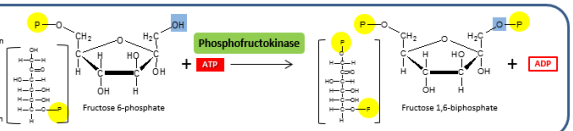
Step 2

A rearrangement of the chemical structure occurs. This is called isomerization and results in the movement of the carbonyl oxygen (C=O) from carbon 1 to carbon 2. This reaction is reversible.



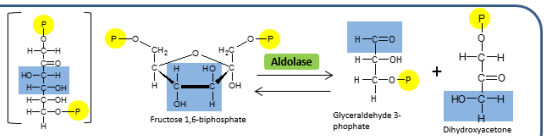
Step 3

The hydroxyl group (OH) on carbon 1 of fructose 6-phosphate is phosphorylated by ATP. This makes fructose 1,6-biphosphate. This is another step that requires energy in the form of ATP



Step 4

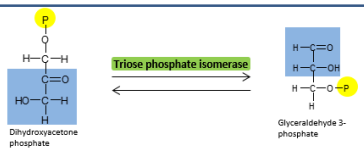
The six-carbon sugar called furctose that has been phosphorylated twice is cleaved to produce two 3 carbon molecules. Only Glyceraldehyde 3-phosphate can continue through the remaining steps of glycolysis.



Step 5

The other product from step 4 is called Dihydroxyacetone phosphate.

An isomerization reaction occurs to rearrange the chemical structure and create Glyeraldehyde 3-phosphate. Because of step 5, each of the products from step 4 ultimately continue through glycolysis as Glyceraldehyde 3 phosphate .

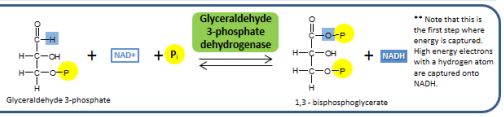


10 STEPS OF GLYCOLYSIS... CONT

****Don't forget that steps 6-10 happen TWICE. Once for each 3 carbon molecule generated in step 5****

Step 6

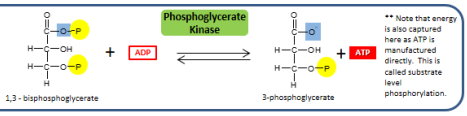
The aldehyde group of Glyceraldehyde 3-phosphate is dehydrogenated. The hydrogen with 2 electrons becomes part of NAD+ (becoming NADH). A phosphate is bonded at the number 1 carbon to make 1,3 -bisphosphoglycerate.



** Note that this is the first step where energy is captured. High energy electrons with a hydrogen atom are captured onto NADH.

Step 7

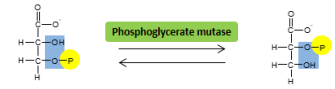
A high energy phosphate from carbon 1 is transferred to ADP to form ATP. This is the first example of "substrate level phosphorylation".



** Note that energy is also captured here as ATP is manufactured directly. This is called substrate level phosphorylation.

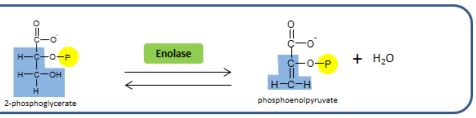
Step 8

In preparation for step 09, the phosphate ester linkage at carbon 3 is removed and relocated at the number 2 carbon.



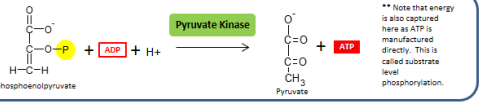
Step 9

A water molecule is removed from 2-phophoglycerate and as a result a double bond is formed between carbon 2 and 3. This creates a higher energy phosphate linkage.



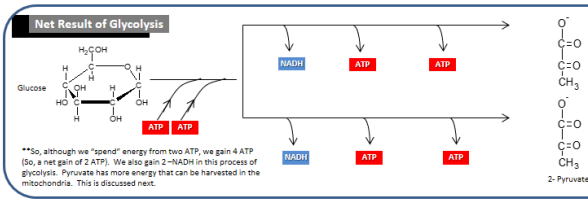
Step 10

A phosphate group is transferred from the carbon 2 of phophenolpyruvate to ADP. This is another example of substrate level phosphorylation. The result is the final product of glycolysis called Pyruvate.



** Note that energy is also captured here as ATP is manufactured directly. This is called substrate level phosphorylation.

Net Result of Glycolysis

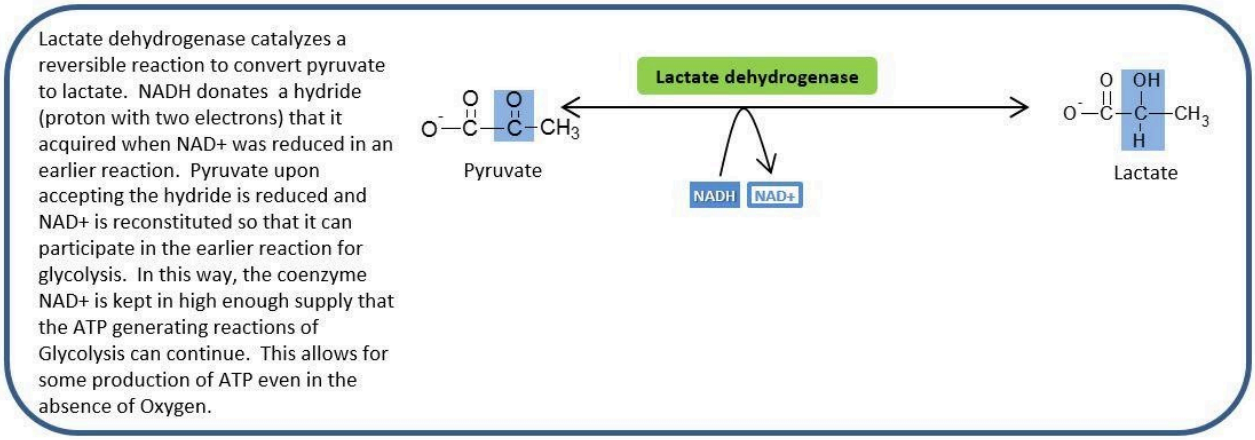


**So, although we "spend" energy from two ATP, we gain 4 ATP (So, a net gain of 2 ATP). We also gain 2 –NADH in this process of glycolysis. Pyruvate has more energy that can be harvested in the mitochondria. This is discussed next.

Anaerobic and Aerobic Use of Pyruvate

Anaerobic

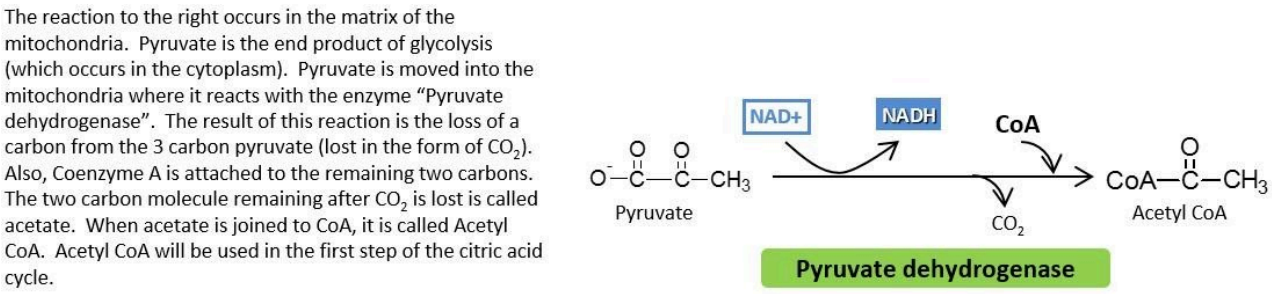
The last step of glycolysis leaves us with two 3-carbon molecules, called pyruvate. The fate of pyruvate depends on the availability of oxygen. If oxygen is available, then pyruvate is shuttled into the mitochondria and continues through several more biochemical reactions called the "Citric Acid Cycle." This is called **aerobic metabolism**. If oxygen is not available in sufficient quantity to the cell, then pyruvate goes through a reduction reaction that results in the production of Lactate (see image below). This is called **anaerobic metabolism**.



Anaerobic Metabolism: Pyruvate Reduction to Lactase. *Image created by JS at BYU-Idaho Fall 2013.*

Aerobic

When there is enough oxygen available to the cell, pyruvate crosses the mitochondrial membrane and is quickly converted to Acetyl CoA. Acetyl CoA enters the Citric Acid Cycle where CoA is removed and the acetate is added to a 4 carbon molecule to make a 6 carbon molecule called “Citric Acid.” As the biochemical steps of the Citric Acid Cycle continue, 2 more carbons are lost as CO_2 and so ultimately all the carbons of pyruvate are lost as CO_2. After 2 pyruvates complete the citric acid cycle, all the carbons of the original Glucose molecule have been released as CO_2.

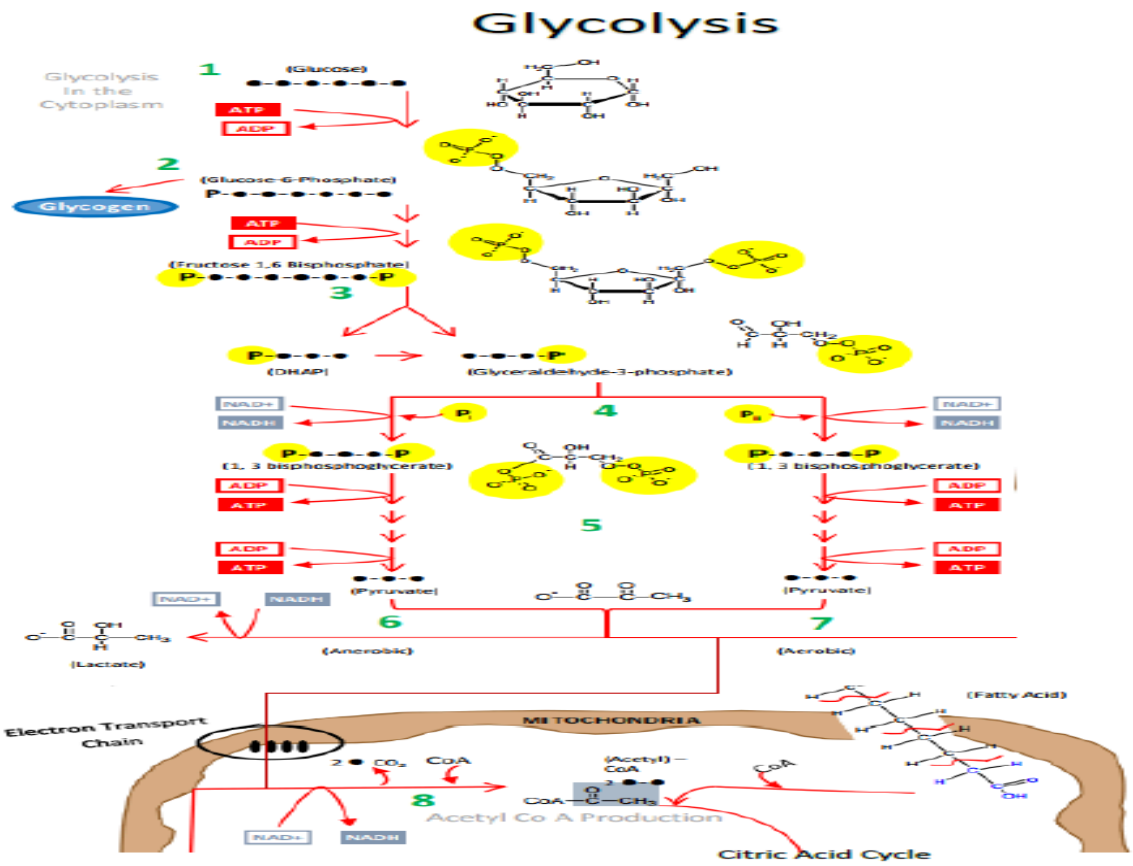


Conversion of Pyruvate to Acetyl CoA. *Image created by JS at BYU Idaho F2013.*

The image above shows the conversion of Pyruvate to Acetyl CoA occurs in the mitochondria and results in the loss of a Carbon as CO_2 and the creation of Acetyl CoA.

Metabolism Summary Part 1: Glycolysis

Below is an image of the process of Glycolysis magnified from the Metabolism Summary image you saw in 8.1. A summary follows for the process of Glycolysis that you have just read about. The green numbers in the image correlate with each of the steps listed below:



Glycolysis, from the "Big Picture" of Metabolism:
Glycolysis, Citric Acid (Krebs) Cycle, Electron Transport Chain, Beta Oxidation and Lipolysis.
Image created at BYU-Idaho by JS 2010

1 Glucose enters a cell and is quickly phosphorylated (meaning a phosphate group is added to the glucose molecule) on the 6th carbon by ATP. This "traps" the glucose in the cell as the charged phosphate group changes the way glucose fits in a glucose transport protein (GLUT). Glucose with a phosphate attached is too large and polar to escape by passive diffusion through a bi-lipid membrane layer.

2 If the enzyme "glycogen synthase" is available, and the cell has enough energy that it does not necessarily need the glucose to make ATP, then this newly phosphorylated glucose may be attached to a chain of glucose molecules called **glycogen**. This is handy because later, when blood sugar begins to drop, glucose will be cleaved from glycogen and made available to go through glycolysis. Also, the phosphate may be removed and the glucose will be put back in the blood to bolster blood sugar levels. These processes called **glycogenesis** (glycogen synthesis) and **glycogenolysis** (glycogen break down) occur in muscle cells to a small extent and in liver cells to a large extent.

3 A phosphorylated glucose that does not become part of the stored glycogen will undergo a conformational change and become fructose. The fructose molecule has another phosphate attached to it from ATP. This double phosphorylated 6 carbon fructose is now primed to be divided into two 3-carbon sugars – each with one phosphate attached. The remaining glycolytic reactions will now happen twice because there are two 3-carbon molecules called Glyceraldehyde-3-phosphate.

4 The energy available in the glucose molecule is found in the form of "chemical energy". This energy exists in the C-H bonds – or more specifically within the electrons that constitute the carbon – hydrogen bonds. The dehydrogenase enzyme in step 4 will remove two hydrogens (2 protons and 2 electrons) from Glyceraldehyde-3-phosphate. Oxidized **Nicotinamide Adenine Dinucleotide (NAD^+)** accepts and bonds with one of the protons and both of the electrons. The

other proton does not bond with the NAD^+ but will be found nearby. This may be written as $:H^- + H^+ + NAD^+ \rightarrow NADH + H^+$.

Because NAD^+ acquires 2 new electrons, we say that NAD^+ is reduced. The 3 carbon molecule that the protons and electrons were removed from is oxidized. This is an example of a redox reaction. For simplification, the reduced form of NAD^+ will be referred to as NADH (instead of NADH + H^+).

Think of NAD^+ as an electron carrier. It is like an empty taxi cab. It comes in and parks near the "dehydrogenase" enzyme and as the reaction occurs, NAD^+ acquires 2 high energy electrons and a proton as passengers. This "taxi" is now occupied and will be referred to as NADH. Later, we will see that these new "passengers" will need to be dropped off for other metabolic reactions to proceed. When NADH unloads its "passengers" NAD^+ is reconstituted and becomes available to go back and participate in reactions again. Without NAD^+ involvement, the dehydrogenase enzyme would not be able to complete the reaction and glycolysis would stop at this point. Notice that if glycolysis stopped at this point, ATP would not be generated in glycolysis because the ATP generation steps are yet to come. It is important to have enough NAD^+ around to keep the reactions going.

Another important effect of the dehydrogenase reaction in step 4 is that an inorganic phosphate (Pi) ends up being bonded to the 3 carbon molecule from step 3. We now have two 3 carbon molecules called 1,3 bisphosphoglycerate.

5 In step 5 there are several biochemical reactions that ultimately accomplish one very important outcome – **Substrate-Level Phosphorylation**. In glycolysis, Substrate-Level Phosphorylation is the transfer of a phosphate group from a 3 carbon organic molecule to ADP. This reconstitutes ATP which can be used in other important energy consuming processes of the cell. Substrate-Level Phosphorylation is different from Oxidative Phosphorylation which will be discussed in Step 12.

Notice that because there are two 3-carbon molecules to donate phosphate groups, 4 ATP molecules will be generated. Notice that for every glucose molecule in glycolysis, 4 ATP are made. However, 2 ATP are required at the beginning steps of glycolysis, so the net production of ATP in glycolysis is 2 new ATP for every glucose molecule.

6 The two 3-carbon molecules left after Substrate-Level Phosphorylation are called pyruvate. **Pyruvate** is the end product of glycolysis. The fate of pyruvate will depend on whether there is enough oxygen available to the cell or not.

If a hypoxic (meaning that oxygen is deficient) condition exists, then a dehydrogenase enzyme will perform a reaction that is actually the reverse of what we saw in step 4. A hydrogen ion and 2 electrons will be removed from NADH and put onto pyruvate. This causes pyruvate to become **lactate**. You have probably heard of lactic acid before. Lactic acid is acid form of the conjugate base lactate. It has often mistakenly been referred to as a negative thing – a waste product or a product that makes muscles fatigue or maybe even a product that causes muscle soreness. While it is true that lactate is produced when muscles work very hard (because the body cannot deliver oxygen fast enough), it is **not** true that lactate causes pain or fatigue. In fact, lactate is easily absorbed and converted back to pyruvate by other cells of the body. Lactate does not last long in the blood and it is **not** something that courses through us like a poison causing all kinds of trouble.

You might be asking why this conversion of pyruvate to lactate is even necessary. Remember that the reactions of step 4 are not possible without NAD^+. If we continually made NADH and had no way to reconstitute or recycle back NAD^+, then we would soon have to stop glycolysis and wait until more NAD^+ became available. Since none of the ATP producing steps of glycolysis can happen until NAD^+ arrives, we would not be making ATP which could kill the cell. Making lactate is a quick way to free up NAD^+ to go back to step 4 and allow the Substrate-Level Phosphorylation reactions to take place. This is called **Anaerobic Metabolism**. Anaerobic metabolism is very fast, but not very efficient (not a lot of ATP per glucose molecule). It is good for sudden bursts of intense activity but cannot sustain activity for very long.

7 If Oxygen is available then pyruvate is transported to the mitochondria. Pyruvate moves across the two mitochondrial membranes and a whole new sequence of metabolic steps proceed in the mitochondrial matrix. The culmination of all

the metabolic reactions in the cytoplasm and the matrix of the mitochondria are called **Aerobic Metabolism**. It is called aerobic because oxygen is used in step 11. Aerobic Metabolism results in much more ATP than were produced by glycolysis alone.

Summary To Be Continued: We will pause our metabolism summary there as we move on to the next stage of metabolism: The Citric Acid Cycle.

8.3

CITRIC ACID CYCLE

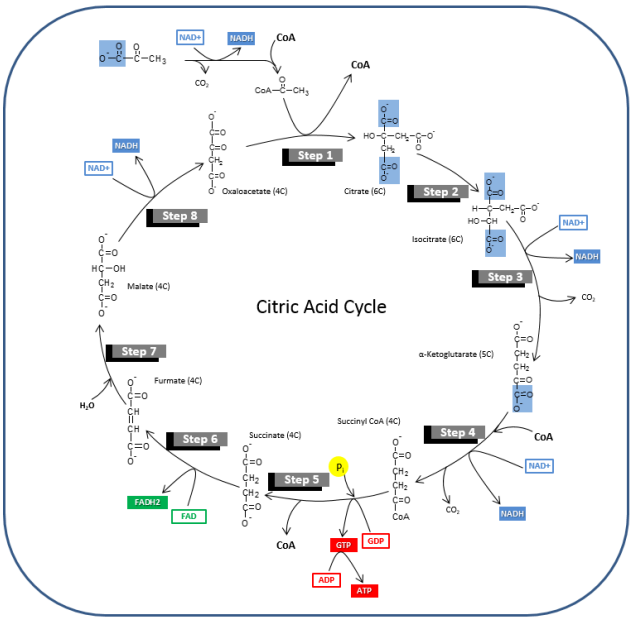
The Citric Acid Cycle is also called the "Tricarboxylic Acid Cycle (TCA) and the "Krebs Cycle." This cycle is a series of biochemical reactions that completes the catabolic pathway for the Glucose molecule that started glycolysis. Energy from the Citric Acid Cycle is captured by electron carriers (NAD and FAD). Also, ATP is generated at one of the steps in this cycle. Remember that you will be expected to understand the citric acid cycle at the level it is shown in the summary figure you have previously downloaded.

After the completion of this phase of metabolism, the following molecules and ATP are made as a byproduct:

Molecule	**Net Yield through Citric Acid Cycle**
ATP	2
NADH	6
FADH2	2

Below is a more detailed figure showing the citric acid cycle. Although it is more than you will need, it may serve as a good reference for you to understand more about the citric acid cycle. It helps to review this figure along with your summary figure to understand more about the citric acid cycle. This figure may help you understand the terminology that you come across in your online research to learn about the citric acid cycle.

FIGURE 9 - 8 STEPS OF THE CITRIC ACID CYCLE

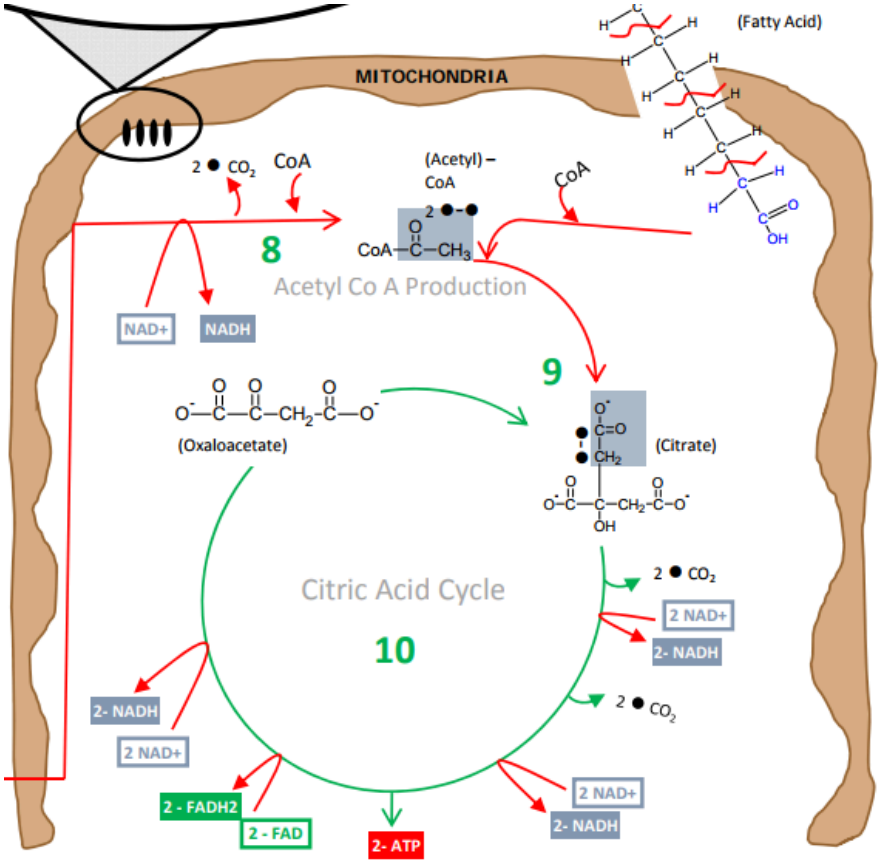


The figure above represents the Citric Acid Cycle (also called the "Krebs Cycle"). The part of the molecule that becomes carbon dioxide is highlighted in a blue box. Notice that a 4 carbon molecule called Oxaloacetate picks up 2 more carbons when it is joined with an acetyl group from Acetyl CoA. Through the beginning steps of the cycle, 2 carbons are lost as carbon dioxide and the molecule is again restored to a 4 carbon state, ready to pick up another acetyl group.

The details of the eight steps above are shown in the following pages. This time, the part of the molecule that undergoes a change is highlighted in blue and the name of the enzyme that catalyzes the reaction is in a green box.

Metabolism Summary Part 2: Citric Acid Cycle

Below is an image of the process of the Citric Acid (Krebs) Cycle magnified from the Metabolism Summary image you saw in 8.1. A continuation of our summary on metabolism follows below for the Citric Acid Cycle up until we reach the Electron Transport Chain where we will return to get the detailed information on this process first before continuing with our summary on all of the processes of metabolism. The green numbers in the image correlate with each of the steps listed below:



Citric Acid (Krebs) Cycle, from the "Big Picture" of Metabolism: Glycolysis, Citric Acid (Krebs) Cycle, Electron Transport Chain, Beta Oxidation and Lipolysis. *Image created at BYU-Idaho by JS 2010*

8 Steps 8-12 complete the story of aerobic metabolism of glucose. After pyruvate is transported into the mitochondria, another dehydrogenase enzyme (actually a very large enzyme complex) will accomplish several things. It will remove 2 protons and 2 electrons from pyruvate. This creates NADH (actually 2 NADHs because there are 2 pyruvates). Also, the reaction results in the loss of a carbon and two oxygen atoms (released as CO_2). Finally, the remaining 2 carbon molecule is attached to Coenzyme A.

Coenzyme A (often referred to as simply CoA) is derived from pantothenic acid (Vitamin B5). **Acetyl CoA** is the name used for the product of the reaction in step 8. The "Acetyl" prefix specifically refers to the 2 carbon group that is being transported by the CoA. Black dots in the summary figure help us keep track of the carbons that originated from the glucose molecule way back at the beginning of glycolysis. Notice that ultimately all the black dots are released as CO_2 so that the metabolism of glucose leaves us with no accumulation of carbons in the cell. Acetyl CoA will enter and participate in the reactions of the Citric Acid Cycle.

9 Step 9 represents the activities of the **Citric Acid (or Krebs) Cycle**. The Citric Acid Cycle involves a lot of steps. The intent for our level of understanding will not involve details of all the reactions.

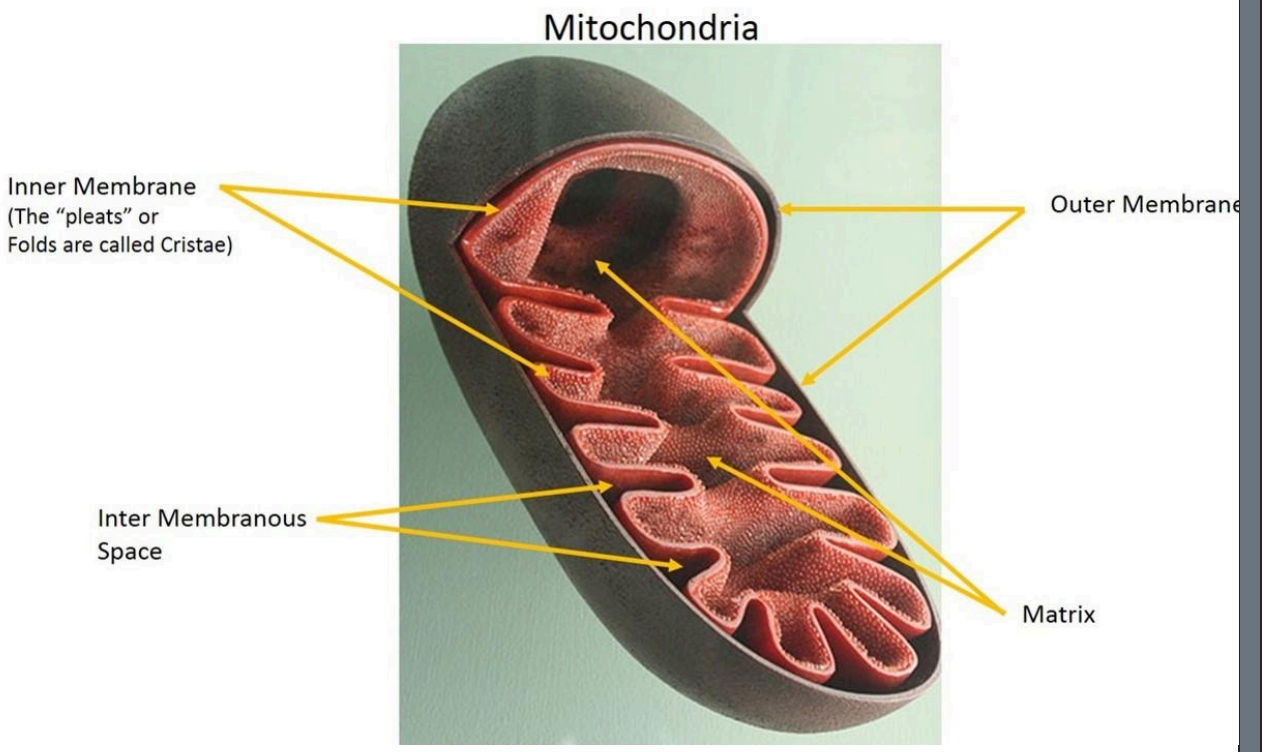
10 The important things to remember about the Citric Acid Cycle are…

1. A 4 carbon molecule called oxaloacetate combines with the acetyl (2 carbon) group of Acetyl CoA (which came from glucose or fatty acids or possibly even some of the amino acids). This will yield a 6 carbon molecule called citric acid. Citric acid will be changed and manipulated as this 6 carbon molecule ends up recycled back to oxaloacetate – thus the term “Citric Acid Cycle”.
2. During the reactions of the citric acid cycle, CO_2 will be lost twice. This means that if you are counting, you will realize that every carbon of the original Glucose or Fat is ultimately lost as CO_2. This is the reason we have to breathe continuously as we do.
3. Hydrogens with electrons are transferred to NAD^+. This creates 6 NADH molecules.
4. FAD is reduced to $FADH_2$. This yields 2 $FADH_2$ molecules.
5. Substrate-Level Phosphorylation will yield an ATP for each turn of the Citric Acid Cycle (or 2 total for each glucose).

8.4

ELECTRON TRANSPORT CHAIN

The Electron Transport Chain is responsible for the synthesis of **most** of the ATP in our body. In order to understand how the electron transport chain works, it is critical that you have a good understanding of what the mitochondria are and how it is organized.



Mitochondria, a double membrane organelle inside the cell. *Image derived from File: Überseemuseum Bremen 200 237.JPG; Author: Sterilgutassistentin;Site: https://books.byui.edu/-cgKk;License: GNU General Public License as published by the Free Software Foundation*

The Mitochondria have an inner and an outer membrane. The inner membrane folds in and out on itself and these fold are called Cristae. Cristae increase the total surface area of the inner membrane. The center of the mitochondrion is called the matrix and is analogous to the cytoplasm of a cell. The Electron Transport Chain reactions take place on the inner membrane.

The term, electron transport refers to the proteins on the inner membrane of the mitochondria that will take hydrogen atoms and electrons from NADH and $FADH_2$ and then ultimately use the energy in the electrons to make ATP. Recall th NAD^+ and FAD picked up high energy electrons and hydrogens from C-H bonds in glycolysis (from the cytoplasm) and the citric acid cycle (in the matrix of the mitochondria).

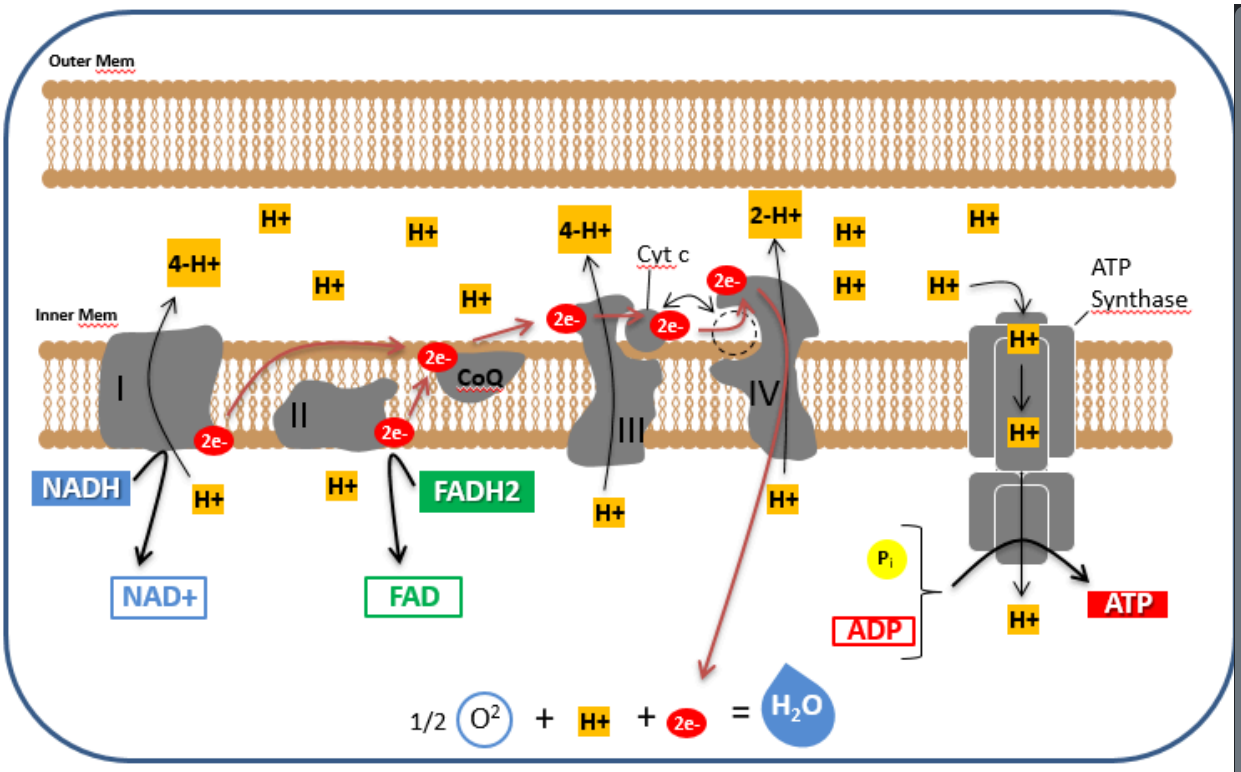
In the inner membrane of the mitochondrion is a series of protein complexes that will receive the electrons and pass them from one complex to another. NADH passes 2 high energy electrons onto a protein complex (**Complex I**) in the inner membrane of the mitochondria. This complex is called *NADH dehydrogenase*. *NADH dehydrogenase* does two things. First, it accepts a pair of high energy electrons from NADH. Second, it uses some of the energy from these electrons to undergo a conformational change. This conformational change is associated with the movement of $4H^+$ ions from the mitochondrial matrix to the intermembranous space (the space between the inner and outer membranes of the mitochondria). Next, these two new electrons on Complex I are moved to Coenzyme Q (CoQ). Coenzyme Q is als called ubiquinone.

$FADH_2$ also passes a pair of high energy electrons to a protein complex (**Complex II**), also called *Succinate dehydrogenase*. Complex II accepts the electrons but does not go through any conformational change that is associated with the movement of H^+ ions. However, Complex II does pass the electrons to **CoQ** just like Complex I did. **CoQ** is a mobile shuttle that moves easily through the membrane and is able to relocate and react with **Complex III**. Complex III has a long name (*Coenzyme Q-Cytochrome c Oxidoreductase*). Complex III also goes by the name *Cytochrome bc1 Complex*. Complex III will undergo a conformational change that is associated with the movement of $4H^+$ ions from the mitochondrial matrix to the intermembranous space. The two electrons are then moved from Complex III to Cytochrome C (**Cyt c**). **Cyt c** another mobile shuttle that is a soluble protein in the intermembranous space that moves easily along the membrane and reacts with **Complex IV**. Complex IV, also called *Cytochrome c Oxidase*, uses some of the electron energy to undergo a conformational change that is associated with the movement of 2 H^+ ions from the mitochondrial matrix to the intermembranous space. Oxygen receives the 2 electrons from Complex IV and reacts with H^+ available in the surrounding fluid to make H_2O or water.

A review of figure 11 below reveals that one NADH results in the movement of 10 H^+ ions from the mitochondrial matr to the intermembranous space. One $FADH_2$ results in the active transport of 6 H^+ ions. The important message in all of this is that electron energy is used to transport H^+ ions to the intermembranous space and this sets up an electrochemical gradient that favors the movement of H^+ ions back into the matrix. This is allowed to happen through another protein complex called ATP synthase. The diffusion of H^+ ions through **ATP synthase** is called "**chemiosmosis**

ATP synthase is made up of two main components. The component found underneath or in the mitochondrial matrix is capable of rotating and also binds ADP, Pi and ATP. ATP synthase also contains channels that allow the diffusion of H^+ ions. As H^+ ions diffuse back into the matrix, ATP synthase is physically rotated and enabled to react with ADP and Pi. There is a resulting phosphorylation of ADP and this yields ATP. This process is called **Oxidative Phosphorylation**.

For each pair of electrons that move from Complex I to Complex IV, about 2.5 ATP can be produced. For each pair of electrons that move from Complex II to Complex IV, about 1.5 ATP can be produced. Therefore, if we round up, it is oft stated that each NADH yields 3 ATP while each $FADH_2$ will yield 2 ATP.



Electron Transport Chain. *Image created by JS at BYU Idaho F2013.*

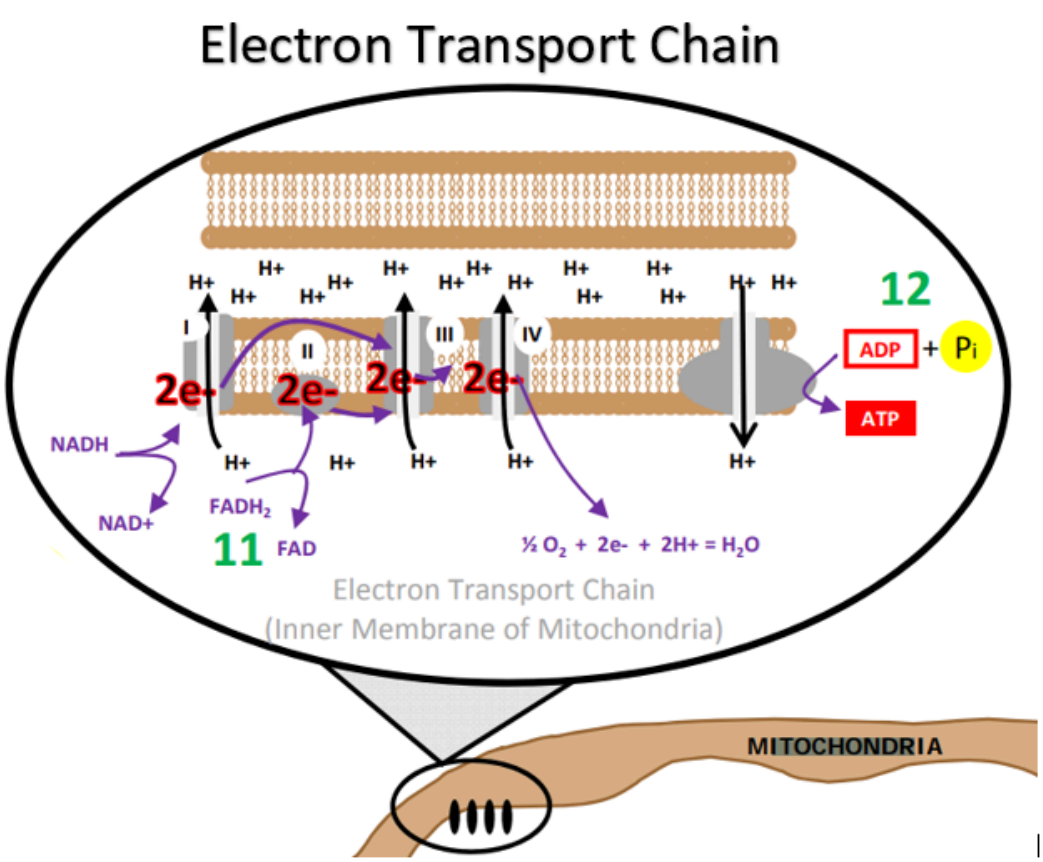
The image above illustrates the Electron Transport Chain. The protein complexes on the inner mitochondrial membran use high energy electrons from NADH and FAD_2 to move H^+ ions to the intermembranous space. The H^+ concentration gradient is then used to make ATP through the enzyme complex called ATP Synthase. Oxygen is the final electron acceptor and becomes water.

A quick recap of what has happened so far might go like this: Electrons and hydrogen ions were harvested from the C-bonds of glucose. These high energy electrons with hydrogen are carried from the reactions of glycolysis and the citric acid cycle to the electron transport chain on the inner membrane of the mitochondria. The electron transport chain takes these high energy electrons and gradually "uses" the energy to pump hydrogen ions into the intermembranous space. As the energy in the electrons is used, the electrons don't have enough energy to form a C-H bond anymore, but they can form an O-H bond. Thus, oxygen comes along and accepts the electrons and hydrogen to form water. The cycle is complete and water can once again be used by a plant somewhere to participate in the photosynthetic reactions that will excite O-H bond electrons again. The hydrogen ions that have been pumped into the intermembranous space are allowed to flow down their electrochemical gradient through ATP synthase. ATP is generated as a result and ATP is used to run the many molecular processes in our cells that keep us healthy and alive. As energy is released in these many reactions of metabolism a little bit is lost as heat. Indeed, metabolism is responsible for a portion of our body heat.

Glucose is not the only molecule with C-H bond energy to use in metabolic reactions. Lipids and Proteins are also metabolized by our cells.

Metabolism Summary Part 3: Electron Transport Chain

We will continue on with our summary of the metabolic process using the Electron Transport Chain process magnified from the “Metabolism Summary” image from section 8.1. (Green numbers from summary correlate with green number on the image below.)



Electron Transport Chain, from the "Big Picture" of Metabolism:Glycolysis, Citric Acid (Krebs) Cycle, Electron Transport Chain, Beta Oxidation and Lipolysis. *Image created at BYU-Idaho by JS 2010*

11 NADH is carrying a proton and 2 high energy electrons that need to be "dropped off". $FADH_2$ is also carrying high energy electrons and a couple of protons. These electron "carriers" are able to donate these electrons to an enzyme complex found in the inner mitochondrial membrane. Think of the "**electron transport chain**" as being like a bucket brigade. A series of proteins pass 2 electrons from one to another. Sometimes when the electrons are passed, a little b of the energy from the electrons is used to induce a conformational change in some of the protein structures. This conformational change results in the transport of protons from the inside of the mitochondria to the intermembranous space (the space between the inner and outer mitochondrial membranes). Also, some of the energy released as the electrons move through the electron transport chain is given off as heat. NADH donates to the electrons to the electro transport chain at complex I and $FADH_2$ donates electrons at complex II. There are more protons pumped from electrons moving down the transport chain from complex I than from complex II. For this reason, NADH yields more ATP ultimately than $FADH_2$. Whether from NADH or $FADH_2$, any donated electrons will move down the transport chain the last protein acceptor and cannot go back to previous components of the chain.

Oxygen accepts the electrons from the last protein complex (complex IV) of the chain. As oxygen accepts the electron the oxygen becomes reactive and capable of forming a covalent bond with two protons and water is formed (H_2O). If you have ever wondered what oxygen does and why it is so important to breathe into our bodies, now you know. Oxyge is the final electron acceptor.

Notice that NADH becomes NAD^+ at the beginning of the electron transport chain. Also, $FADH_2$ becomes FAD. This recycles these electron carriers such that they can be used again in earlier metabolic reactions. This has been mentioned, but it is worth mentioning again. Without NAD^+, reactions that use NAD^+ cannot occur.

Think about this question: *What if there were no oxygen available?* Predict what would happen to the metabolic pathways discussed…

Answer: If there were no oxygen, then the high energy electrons would sit on the protein complexes in the electron transport chain. They can't go back and they can't move forward because there is no FINAL electron acceptor. Comple I would have electrons that couldn't move as well, so NADH could not "drop off" any more electrons. Since NADH coul not become NAD^+, it would not be very long before a severe shortage of NAD^+ stopped the metabolic reactions from occurring. This would basically grind the Citric Acid Cycle to a stop. Also, pyruvate could not become Acetyl CoA. Basically, none of the mitochondrial metabolic processes could occur. Aerobic Metabolism would not be possible. However, Glycolysis would still be possible. You might recall the possibility to regenerate NAD^+ in the cytoplasm if Lactic acid is made (Anaerobic Metabolism).

Hopefully, now you can see why oxygen is so important.

12 In step 11, we learned that as high energy electrons passed down the chain of protein acceptors, energy was used t move H^+ ions into the intermembranous space. This generates a proton gradient. This means that there will be a highe concentration of protons in the intermembranous space than there is inside the mitochondrial matrix. This proton gradient represents "potential energy" because the protons will try to flow down their gradient if a passageway opens u and allows such movement.

Step 12 represents the idea sometimes referred to as **chemiosmosis**. This is a term that refers to the fact that protons tend to flow down their gradient through a selective protein channel. This protein channel, called ATP-Synthase is a ver intricate and specialized molecular machine. This protein literally turns as the protons come through it and this kinetic energy is used to bring ADP and inorganic phosphate together so that ATP is created. The synthesis of ATP through chemiosmosis is referred to as **Oxidative Phosphorylation**.

While glycolysis gives us 2 ATP per glucose molecule, the electron transport chain gives us approximately 34 ATP per glucose molecule. We say "approximately" because it is difficult to say exactly how many ATP we get. This is because some ATP is used to shuttle molecules in and out of the mitochondria and there is likely some "leaking" that occurs when protons from the intermembranous space accidentally escape by some other way than through the ATP synthas enzyme complex. However, it is generally accepted that aerobic metabolism yield between 18 and 19 times more ATP than anaerobic metabolism. It is a good thing that our cells have mitochondria!

Although aerobic metabolism gives us much more ATP, it takes longer to do it. You might think of aerobic metabolism being useful for endurance activities but less useful for activities that require both high speed and high-intensity work.

Molecule	**Conversion Rate to ATP**	**ATP Yield through Electron Transport**
2 NADH from Glycolysis	3 ATP / NADH	6 ATP
2 NADH from Pyruvate Oxidation		6 ATP
6 NADH from Citric Acid Cycle		18 ATP
2 FADH from Citric Acid Cycle	2 ATP /FADH	4 ATP
Net Total from Electron Transport		**34 ATP**
Net Total from Glycolysis + Citric Acid Cycle (Remember 2 ATP used to start glycolysis)		4 - 2 + 2 = 4 ATP
Total ATP from the complete aerobic metabolism of one glucose molecule		**~38 ATP (minus some loss)**

8.5

LIPID AND PROTEIN METABOLISM

8.5.1

Lipid Metabolism

Beta oxidation is the term used to describe a series of reactions that break down a fatty acid into 2 carbon acetyl groups which are associated with Coenzyme A (see figure 12). Each time an acetyl CoA is generated from a fatty acid, the fatty acid re-enters the Beta oxidation biochemical pathways to remove the next 2 carbon fragment. This occurs until the entire fatty acid chain has been broken down in this way. Each time a beta oxidation cycle occurs, NADH and $FADH_2$ are generated. Also, each time an acetyl CoA from beta oxidation goes through the Citric Acid Cycle, 3 NADH, 1 $FADH_2$ and 1 ATP are generated. Since a fatty acid is many carbons long (most often found in lengths of 16 or 18 carbons), many acetyl CoA molecules can be acquired from a triglyceride molecule. Enough ATP is made from all the NADH and $FADH_2$ that it becomes clear that fat molecules give us more ATP per gram than glucose molecules.

Lipogenesis is the term used to describe the process of making new fat. Fatty acid chains can be synthesized by combining Acetyl groups which adds carbons to a growing fatty acid chain. It is almost like Beta oxidation in reverse, but the reactions use different enzymes and occur in a different place. While beta oxidation occurs in the matrix of the mitochondria, lipogenesis occurs in the cytoplasm of cells (mostly in the liver and adipocytes). Cells that synthesize fat have an enzyme complex made up of about 7 protein enzymes called **Fatty Acid Synthase**. When cells have excess glucose, there arises an excess of Acetyl CoA molecules. This upregulates lipogenesis and explains how diets high in sugar can cause increased adipose tissue.

Ketoacidosis is a complication that occurs when the body is not metabolizing sugar. This may occur in times of excessive dieting, fasting, or malnutrition. The most common cause of ketoacidosis is Type I Diabetes. In type I diabetes, there is no endogenous insulin and sugar cannot get into the fat and muscle cells which make up the largest percentage of body tissue by volume and weight. This means that these cells will catabolize predominately fat for ATP production (fat does not require insulin to get into cells). As increasing amounts of fat molecules are broken down through beta oxidation, accumulation of acetate and acetyl CoA may occur as the Citric Acid Cycle reaches a limit on how many acetyl CoA molecules it can take in at the first biochemical step. These two carbon products begin to spontaneously react with each other and produce 4 carbon molecules referred to as ketone bodies. The three most common ketone bodies are **acetone**, **acetoacetate**, and **beta-hydroxybutyrate**. These molecules are acidic and in high quantities can lower the pH of the blood. Also, acetone is volatile and can escape through the lungs and give a particular smell to a person's exhaled breath. The smell has been described as being similar to finger nail polish remover (which contains acetone).

Fatty acids must be "activated" before they are transported into the mitochondria. Activation involves the attachment of Coenzyme (CoA). The result is a fatty acid derivative called Fatty acyl-CoA. Fatty acyl-CoA goes through a series of steps illustrated below. This process is called beta oxidation which suggests that the molecule will be oxidized at the beta carbon and then cleaved to yield Acetyl CoA (last step below). The Acetyl group is highlighted in blue in the figure below. The alpha (α) and beta (ß) carbons are labeled on the fatty acid. Notice that after Acetyl CoA is produced, the α and ß carbons for the next cycle are illustrated in gray. Palmitoleic acid is one of the most common fatty acids in animals and is the fatty acid used in this illustration. However, fatty acids can be any length with the most common ones between 14 and 18 carbons long. Complete beta oxidation of palmitoleic acids yields 8 Acetyl CoA molecules that can metabolize further in the citric acid cycle. The enzymes that catalyze each step are depicted in green boxes.

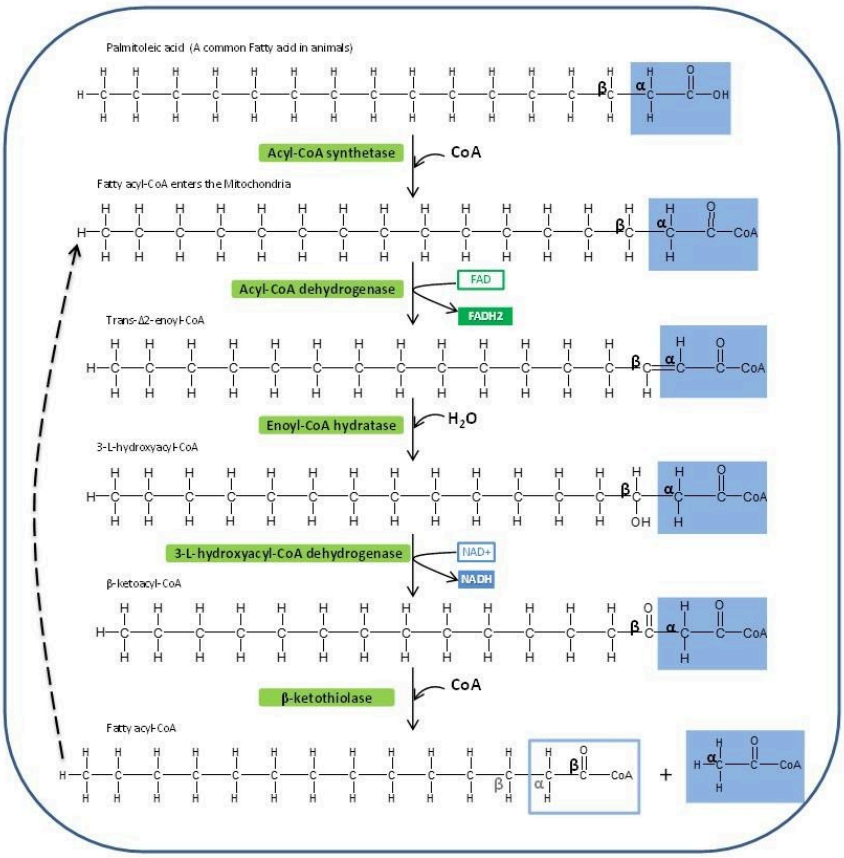
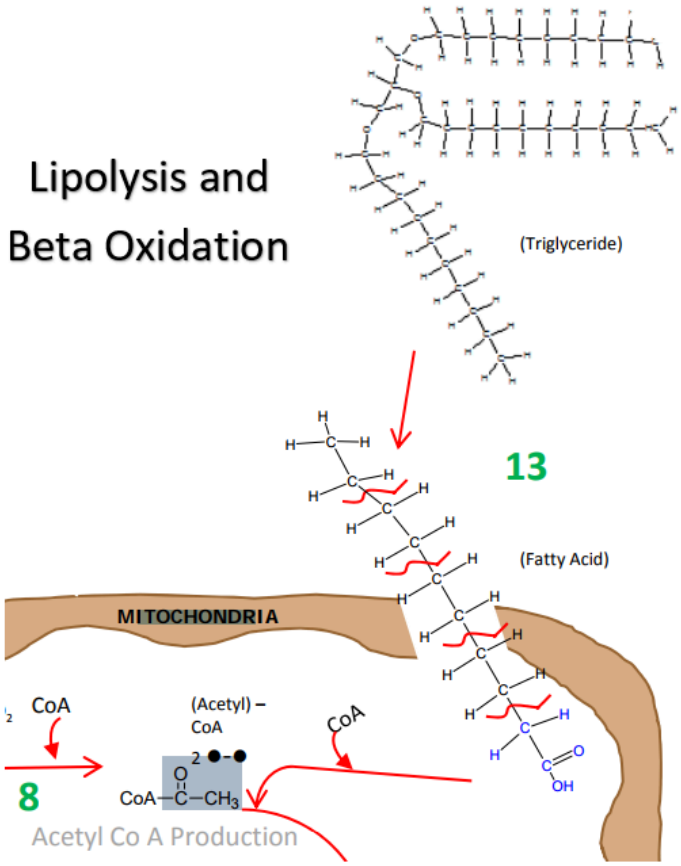


Image created by JS at BYU-Idaho Fall 2013.

Metabolism Summary Part 4: Beta Oxidation and Lipolysis:

Below is the final installment of the Metabolism Summary:



Lipolysis and Beta Oxidation, from the "Big Picture" of Metabolism:Glycolysis, Citric Acid (Krebs) Cycle, Electron Transport Chain, Beta Oxidation and Lipolysis. *Image created at BYU-Idaho by JS 2010*

13 Fat can also be used to make ATP. The metabolism of fat is called Beta Oxidation. Triglycerides are dismantled into glycerol and fatty acids. The glycerol can actually be converted into Glyceraldehyde-3-phosphate and then it completes the glycolytic reactions. The fatty acids are transported into the Mitochondria. Once in the Mitochondrial matrix, fatty acids are dismantled 2 carbons at a time and each 2 carbon piece is converted to Acetyl CoA. Acetyl CoA enters the Citric Acid Cycle. A single triglyceride molecule can ultimately yield a lot more Acetyl CoA than a glucose molecule. So, we say that fat is much more energy dense than sugar because we get more ATP per gram of fat than we do with sugar. However, notice that the end product of Beta oxidation is Acetyl CoA (not pyruvate). So, anaerobic metabolism is not possible with fat metabolism. We can only burn fat when there is enough oxygen available to the cell.

8.5.2

Protein Metabolism

So far this reading has focused on the metabolism of sugars and fats. Indeed, sugars and fats make up the large majority of organic molecules processed as fuel in our cells. However, proteins can be metabolized to make ATP as well. Proteins are responsible for most of the structure and function in our body tissues, so we probably don't want to metabolize them too extensively. In fact, the body employs several regulatory mechanisms to spare body proteins from metabolism. But, in cases of prolonged fasting, or diets high in protein (like our average American diet), we see amino acid metabolism become an important part of ATP synthesis. In fact, the average American diet has much more protein than we need and so we can even see the products of protein metabolism being used to synthesize fatty acids and triglycerides which are stored in our fat cells. When proteins undergo catabolism, they are broken down into individual amino acids. Amino acids differ with respect to the "R group. The "R" group will determine where in the metabolic cycles that the amino acid products will enter. Notice in figure 13 that there are several metabolic entry points for amino acids in the biochemical pathways we have discussed.

Gluconeogenesis

The conversion of pyruvate to acetyl CoA is an irreversible reaction. This means that when fatty acids are metabolized to form acetyl CoA, it is not possible to turn the acetyl CoA back to pyruvate or any earlier glycolytic product. Also, acetyl CoA is 2 carbons long and 2 carbons are lost in the early reactions of the Citric Acid Cycle. For both of these reasons, it is not possible to use fatty acids to make glucose. In order to make glucose from scratch (**Gluconeogenesis**), our cells have to use a substrate that is not acetyl CoA and will not go through CO_2 expelling steps. In the figure below, we see that some amino acids can enter the metabolic pathways in places that meet these requirements. Therefore, amino acids are the best choice for a raw material to make glucose. When amino acids enter the metabolic pathways for the purpose of making glucose, the reactions of glycolysis more or less runs in reverse to synthesize a new glucose molecule. The liver is particularly good at doing this. Gluconeogenesis is stimulated by hormones in the body that are released when blood sugars become low.

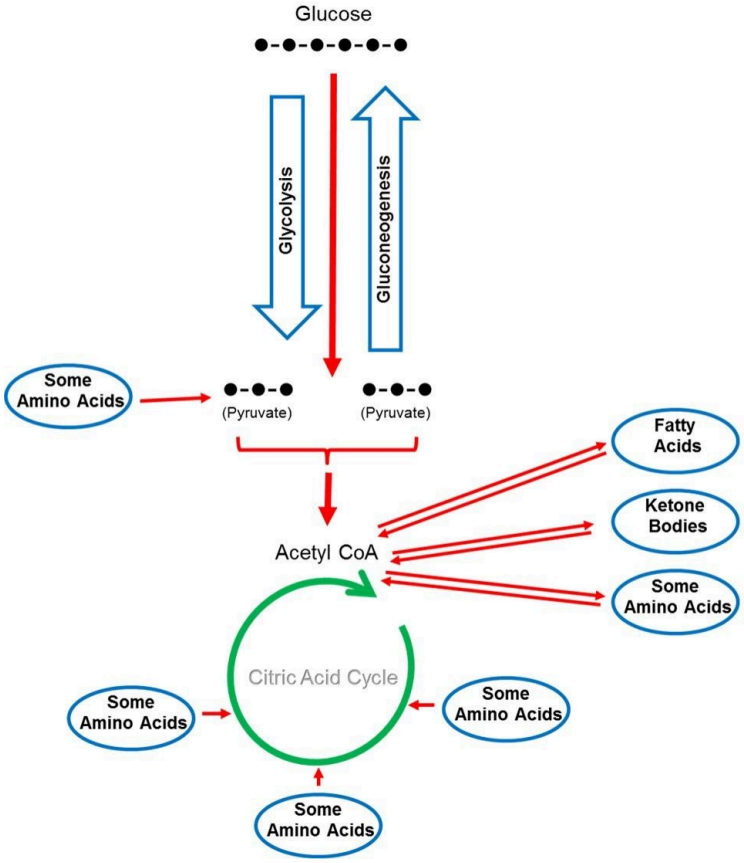


*Image created by JS at BYU-Idaho Fall 2013.***This illustration shows the metabolic entry point of carbohydrates, fatty acids, and amino acids (from proteins). Notice that many of the reactions are reversible.**

9.0

MODULE 9: CONTROL OF BODY MOVEMENT

9.1

VOLUNTARY AND REFLEXIVE CONTROL OF MUSCLES

It is a beautiful, albeit cold, winter evening in Rexburg, Idaho. It just snowed six inches, and your FHE group decided that you could not pass up the opportunity to go sledding at the sand dunes. On the first run of the night, your best friend gets huge air on an unseen jump and then lands awkwardly on her back. You quickly sled down to check on her, careful to avoid the jump. When you reach her, she is sitting up but looks confused. You ask her if she is ok and if she remembers what happened, but she just looks at you. Then, without warning, she turns and vomits on the ground. You get the others and decide to take her to the emergency room to get checked out. She is a little wobbly on her feet at first but is able to walk to the car on her own. After waiting in the exam room for a time, the doctor comes in and asks her a series of questions. He then pulls out a mini flashlight and shines it in her eyes one at a time. Curious about why he is shining a light in her eyes, you ask him. He responds that he is evaluating her brain by checking her reflexes. He then has her stand up and walk across the room as he observes. After finishing his exam, he turns to you and says that she has a mild concussion, but she should be fine. He gives you some instructions and warnings, and then you take her home. As you leave, you wonder:

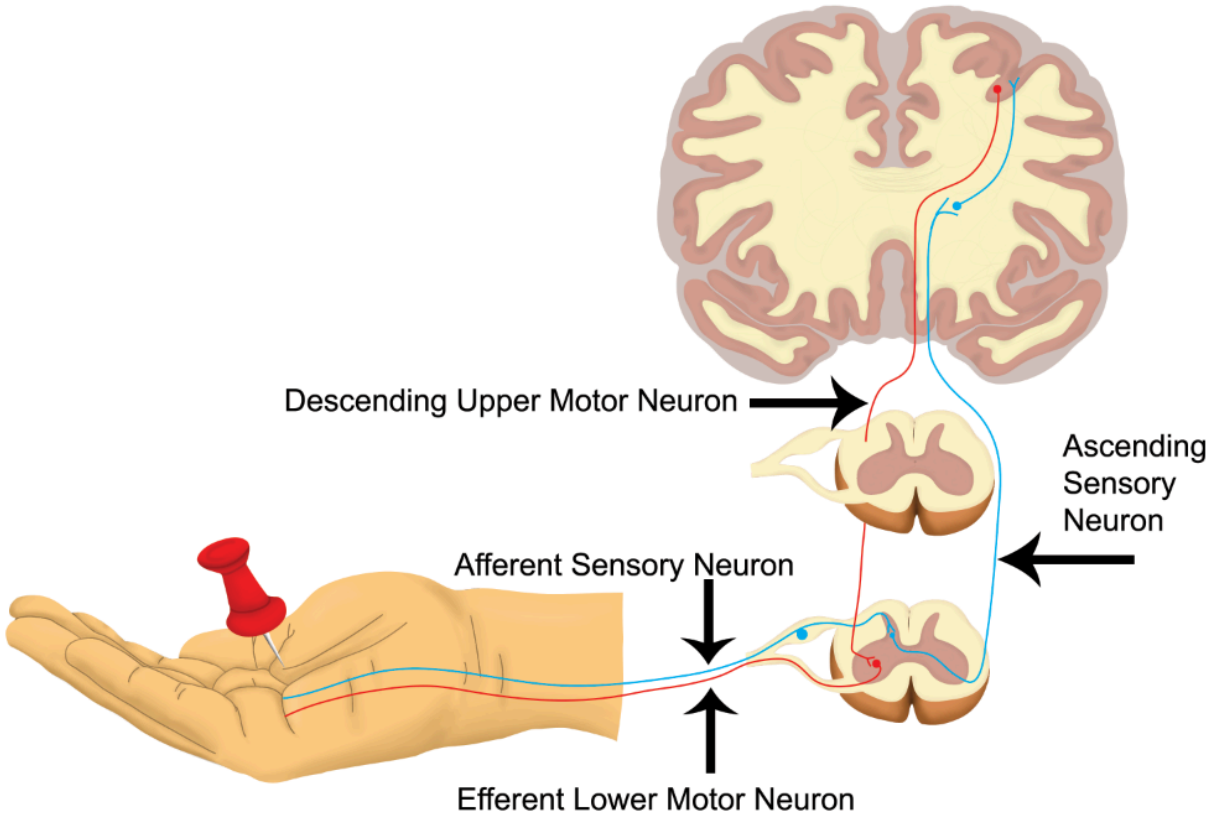
What are reflexes? (You used to think the only way to test reflexes was to hit someone's knee with a hammer!) How do reflexes work? How could the doctor tell that your friend was ok simply by looking in her eyes? What was he looking for as she was walking? How is the nervous system able to control both conscious and unconscious body movements? These are some of the questions that we will attempt to answer in this unit.

9.1.1

Voluntary Control of Muscles

In order to understand reflexes and unconscious movement, we must first examine how voluntary movements are controlled. Voluntary movements, such as walking upright, are rather complex involving multiple areas of the central nervous system (CNS) and peripheral nervous system (PNS). It is no wonder that it takes time to learn to walk or ride a bike. Once learned, these movements are consciously initiated and then carried out almost automatically. Why are we taught that practice makes perfect? It is because the more we practice a skill the more automatic it becomes. We commonly refer to this phenomenon as "muscle memory."

Such movements depend on **upper motor neurons** (UMN) and **lower motor neurons** (LMN). The cell bodies of upper motor neurons are found in the cerebral cortex, where planning, initiation, and coordination of movement occur. Upper motor neurons then synapse with lower motor neurons in cranial nerve nuclei or in the anterior horn of the spinal cord. The lower motor neurons then leave the CNS and synapse with skeletal muscle at the neuromuscular junction. It is this single neuron system from the spinal cord to the muscle that we refer to as the somatic nervous system. To summarize, upper motor neurons initiate movement by sending impulses to lower motor neurons which then relay that information to the skeletal muscle. Thus you can say that voluntary movement comes from the top down and reflexes come from the bottom up. The synapse between the upper motor neuron and the lower motor neuron in the spinal cord is where modulation of both voluntary and reflexive movement takes place.



Withdrawl Reflex: An example of Involuntary Movement that uses sensory and motor neurons of the central and peripheral nervous system. *Image by BYU-I student Becky T. 2018*

The lower motor neuron and each skeletal muscle that it innervates is called a **motor unit**. This has been discussed before, but a quick review would be as follows. A motor unit is a single motor neuron and all the muscle fibers that it innervates. Every time that a motor neuron sends an action potential it will cause an action potential in each of the muscles that it supplies. The size of motor units is quite variable. In areas where we need more precise control each motor neuron innervates very few muscle fibers and in other areas where we need to generate power and precision is not as vital each motor neuron innervates many muscle fibers. Therefore in the places like the eyes and fingers, we would have a low muscle fiber to motor neuron ratio whereas in the large muscles of the legs we would have a higher muscle fiber to neuron ratio. The thigh muscles often have a thousand or so fibers per motor unit, while the delicate muscles that move the hand or control eye movement may have only three to five muscle cells per motor unit. The strength of muscle contractions can vary from weak to very strong. For instance, picking up a feather doesn't require much effort (less overall motor units or using motor units with fewer muscle cells per unit). But, lifting a car would require a much stronger contraction (more motor units and use of units with more muscle cells per unit).

Because muscle fibers contract in an all-or-none fashion the main mechanism for increasing the force of contraction is to stimulate more muscle fibers. This process of increasing the number of motor units activated is called recruitment.

Before a muscle fires, there are several regions of the brain, including the cerebral cortex, basal nuclei (basal ganglia), and cerebellum that work together to control and facilitate the desired movement. We can break this process down into three basic steps: 1) planning, 2) initiation, and 3) execution. The first two steps are mainly controlled by different areas of the cerebral cortex and the last step involves relaying the command from the CNS through the PNS to the muscles involved in the movement.

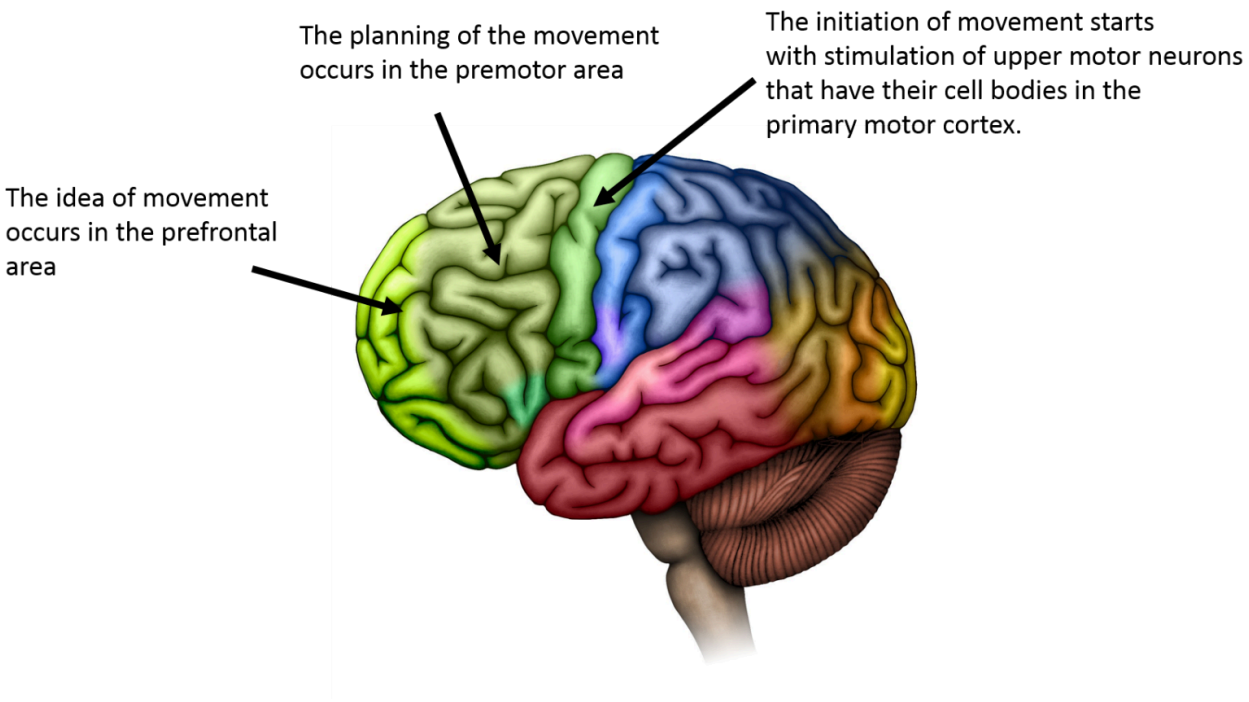


Image drawn by BYU-Idaho student 2014

The planning step includes forming an idea of what you want to do in the pre-frontal or motor association area and then organizing and coordinating the sequence of events in order to accomplish the movement, which takes place in the premotor area. In the initiation step, action potentials are sent to the upper motor neurons of the primary motor area in the precentral gyrus, which initiates the movement. This signal is sent via action potentials to the lower motor neurons in the cranial nerve nuclei of the brainstem or the anterior horn of the spinal cord. These signals travel from the cerebral cortex down to the brain stem and spinal cord and are thus referred to as descending tracts. The largest of the descending tracts that control voluntary movement is the lateral corticospinal tract. You have probably heard that the right side of your brain controls the left side of your body and that the left side of your brain controls the right side of your body. This is due to the fact that the upper motor neurons from the lateral corticospinal tract cross over from the right side of the brain to the left side of the spinal cord and then to the left body extremities. This crossing over, called decussation, occurs in a region of the brainstem called the medullary pyramids. The final step is the execution of the movement. This takes place as the LMNs send excitatory action potentials to the skeletal muscles responsible for the desired movement, called agonist muscles. They simultaneously send inhibitory action potentials to the antagonist muscles that would oppose the desired movement.

This three-step process that we just explained is much more complex than described above. At each step, we are also receiving input from sensory receptors called **proprioceptors**. They relay information about the position of our body and extremities at any given moment. They also tell us about the direction and rate at which we are moving. We also receive sensory input from our eyes, ears and other sensory organs that we use to modify our movements and react to a dynamic environment. These reactions must take place very quickly to allow adjustments in real time and prevent injury, and thus they function somewhat independently of the higher brain areas. We refer to these reactions as reflexes.

9.1.2

Reflexes

Recall from module one the importance of homeostasis. Reflexes are the mechanism whereby the body is able to sense changes and respond appropriately in order to maintain homeostasis. Reflexes take place automatically and unconsciously for good reason. Can you imagine what your life would be like if you had to think about the temperature of the room you just entered and then make conscious adjustments in order to maintain your internal temperature within the normal range? In order to understand how reflexes work, we will examine the basic components of a reflex arc and how they interact to generate the desired effect.

Reflexes can be classified in many different ways, but we will examine them as either somatic: involving control of skeletal muscle, or autonomic: involving control of smooth muscle, cardiac muscle, glands etc. Regardless of their classification, all reflexes have 5 basic components:

1) A **receptor** that detects change. Here are the different types of receptors we see in the body:

- A **mechanoreceptor** is a sensory receptor that responds to mechanical pressure or distortion. Touch, pressure, stretching, sound waves, and motion can all activate mechanoreceptors.
- A **chemoreceptor** detects certain chemical stimuli in the environment. For example, chemoreceptors in the carotid artery are sensitive to the partial pressure of carbon dioxide in the blood; they signal the respiratory center in the brain to increase or decrease the rate of breathing. The sensations of smell and taste happen because of chemoreceptors located in the sensory organs of your body.
- **Thermoreceptors** are specialized nerve cells that are able to detect differences in temperature.
- **Photoreceptors** are the cells in the retina that respond to light. These receptors convert light into signals that give us our vision.
- **Nociceptors** are sensory receptors that detect signals from damaged tissue or the threat of damage and indirectly also respond to chemicals released from the damaged tissue by sending signals to the spinal cord and brain. Nociceptors conduct signals that are generally perceived as pain.

2) An **afferent** or **sensory neuron** that relays information.

3) A **control center** that evaluates the incoming information and determines an appropriate response.

4) an **efferent** or **motor neuron** that carries information away from the control center.

5) An **effector** where the action is carried out.

These components form a basic circuit that is called a reflex arc. The control center portion of the reflex arc can be quite simple (as in the synapse of an afferent neuron onto an efferent neuron in a stretch reflex), but it is often more complex involving one or more interneurons. Simple reflexes involving a single synapse between two neurons are termed **monosynaptic**, whereas reflexes involving two or more synapses are termed **polysynaptic**. We will examine several types of somatic and autonomic reflexes because of their clinical significance. However, please keep in mind that this is not an exhaustive list, and there are many other significant reflexes.

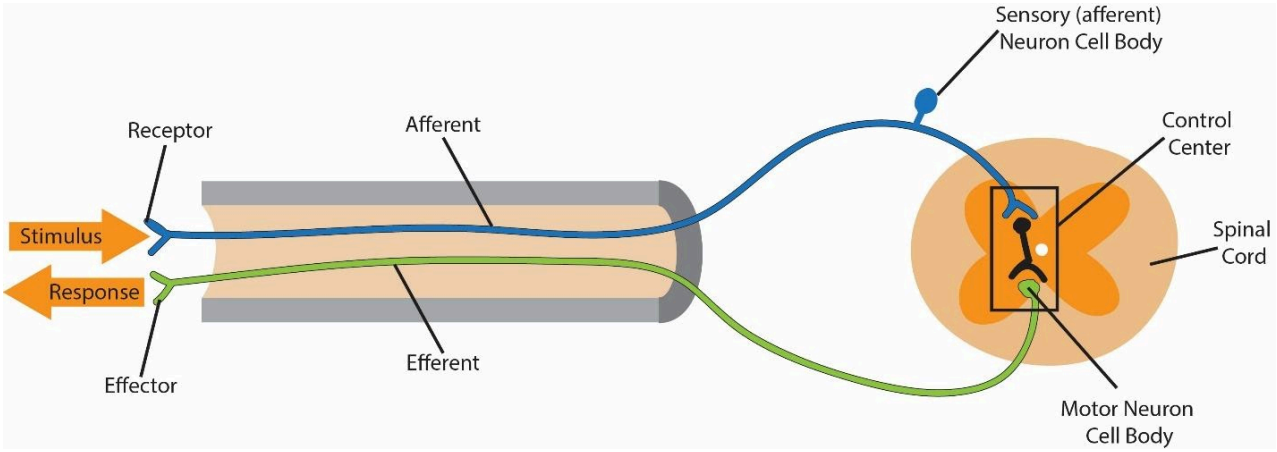


Image created by JS at BYU Idaho Fall 2013.

Somatic Reflexes

In our discussion, we will examine four major reflexes that are integrated within the spinal cord: the stretch reflex, the Golgi tendon reflex, the withdrawal reflex and the crossed extensor reflex. Although each of these reflexes is integrated within the spinal cord, they can be influenced or modified by higher brain centers to either exaggerate or suppress the response. Somatic reflexes involve specialized sensory receptors called **proprioceptors** that monitor the position of our limbs in space, body movement, and the amount of strain on our musculoskeletal system. The effectors involved in these reflexes are located within skeletal muscle.

Stretch Reflex

Think back to the last time you had a sports physical or a routine physical examination. Why did the doctor tap your leg just below the knee? What information can he possibly gather from this simple procedure? The magic of examining reflexes comes from the phenomenon that, under normal circumstances, a specific stimulus will elicit a predictable response. In the case of the knee-jerk reflex, the expected response is an extension of the leg at the knee. If the reflex is greater than expected (hyperactive), less than expected (hypoactive) or totally absent, that suggests potential pathology. Now let's look at how the stretch reflex works.

Muscle spindles are specialized proprioceptors that monitor muscle length. They are bundles of modified skeletal muscle fibers with extensive sensory and motor innervation. These fibers, called **intrafusal fibers**, run parallel to the contractile skeletal muscle fibers called **extrafusal fibers** that make up the bulk of skeletal muscle. Muscle spindles are scattered throughout skeletal muscle, but they occur in the highest density near tendinous insertions and in muscles involved in fine motor control (i.e. the small muscles of the hand etc). Intrafusal fibers are only capable of contraction at their tapered ends where they are innervated by gamma motor neurons. (The contraction is too weak to contribute to gross movement but is important in maintaining the sensitivity of the muscle spindle while the muscle is either shortened or lengthened.) Sensory neurons innervate the noncontractile central region of the intrafusal fibers. If stretched, the sensory fiber associated with the muscle spindle will be activated and result in stimulation of an alpha motor neuron (a type of lower motor neuron) in the anterior horn of the spinal cord. The alpha motor neurons directly innervate the skeletal muscle where the muscle spindle is located. This is an example of a monosynaptic reflex because the sensory neuron synapses directly with the motor neuron and occurs without any input from the upper motor neuron.

Imagine stepping out of the driver's seat of your car onto a patch of ice in the parking lot. As your weight transfers to your left foot and starts to slide out from under you, what happens? The muscle spindles in your left inner thigh (adductors) are quickly stretched and send a message to your alpha motor neurons in the spinal cord begging for help. The alpha motor neurons then cause contraction of the same inner thigh muscles (adductors) that were stretched, and you narrowly avoid the pain of a groin injury. All of this happens so fast (signals are sent at speeds around 350 miles per hour) that you have already recovered by the time you are aware that you were in trouble. When a muscle is stretched, the muscle spindles are stimulated and thus increase the frequency of action potentials sent to the lower motor

neurons in the CNS. The increased action potential frequency causes alpha motor neurons to rapidly fire, resulting in muscle shortening. This reflexive contraction, in the direction directly opposite to the initial stretch, protects skeletal muscle from damage due to overstretching.

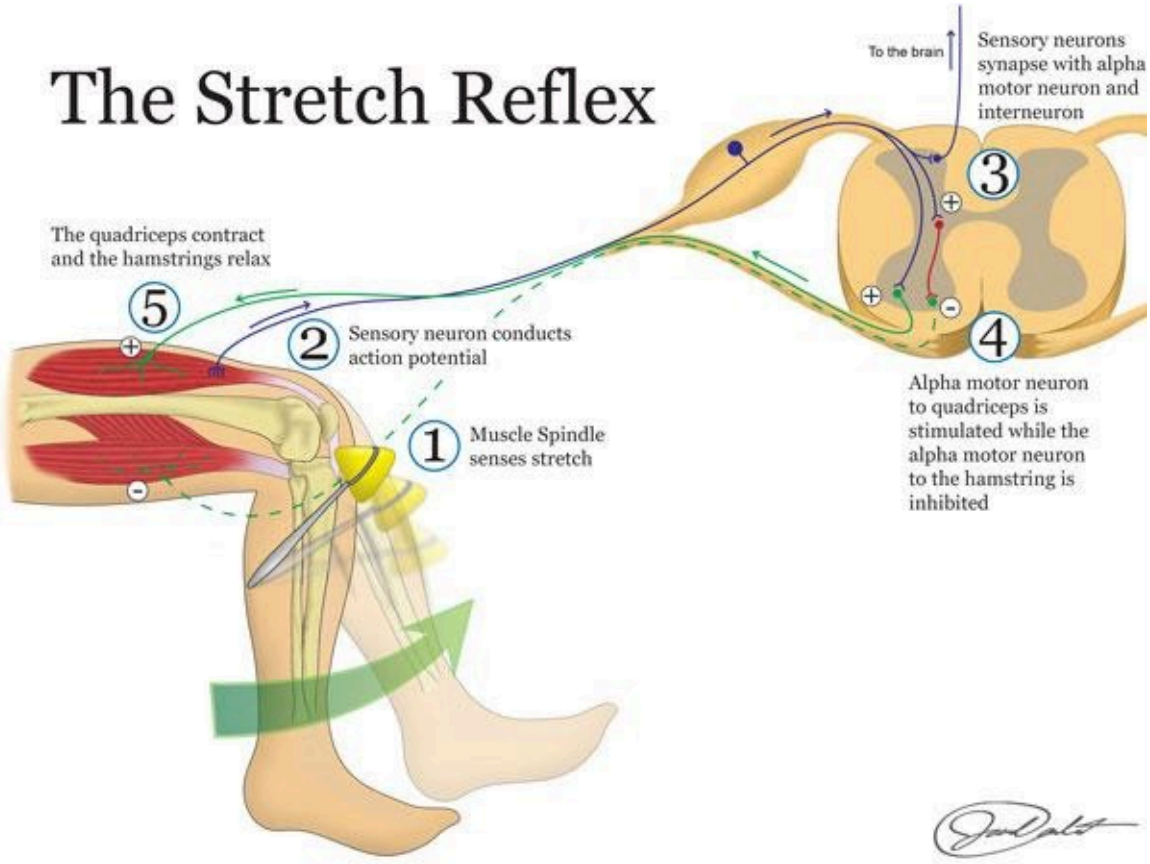


Image drawn by BYU-Idaho student Jared Cardinet Winter 2015

The same process that we described above also relates to other very common situations. For example, as you are reading this you may be experiencing some drowsiness. We will assume that is because you have stayed up way too late! As you get tired you may have experienced the feeling of nodding off, where your head starts to fall forward followed by an almost violent jerking motion as you bring your head upright again. Your muscle spindles are key in maintaining posture, whether we are talking about nodding off in class or whether we are talking about staying upright as you walk down the street.

So, now that the muscle that was being stretched is shortened, what happens to the muscle spindle? Does it become insensitive to further changes in that muscle's length? Remember, we said that gamma motor neurons innervate the contractile ends of the muscle spindle. As the alpha motor neurons activate extrafusal fibers, causing shortening of the muscle, gamma motor neurons activate the muscle spindle. We refer to this as alpha-gamma co-activation. This causes the tapered ends to contract, thus maintaining a baseline tension on the central region of the muscle spindle that is sensitive to stretch. It is in this manner that the muscle spindle is able to maintain its sensitivity through a wide range of muscle length.

In fact, even when a muscle is at rest the muscle spindle sends out a relatively steady stream of action potentials which helps to maintain a low level of muscle activity. This constant tension of the muscle is what we refer to as **muscle tone**.

Up to this point, we have only addressed activation of the muscle group that is being stretched. This is important but body movement is controlled by opposing muscle groups, the agonist and antagonist muscles. The agonist muscle is the muscle that contracts to cause a certain movement to happen and the antagonist is the muscle group that would do

the opposite action. In the example of the knee jerk reflex, the quadriceps would be the agonist and the hamstring would be the antagonist. In order to extend the leg at the knee, we must contract the quadriceps, which we do via activation of the alpha motor neurons, but we must also relax, or inhibit, the hamstring. We accomplish this through a phenomenon called **reciprocal inhibition**. The sensory neuron that synapses with and excites alpha motor neurons supplying the quadriceps also synapses with an inhibitory interneuron. The inhibitory interneuron effectively shuts down the alpha motor neurons to the hamstring. This allows the leg to extend at the knee.

Golgi Tendon Organ (GTO)

Whereas muscle spindles respond to stretch another type of sensory system responds to tension. You might think that stretch and tension are pretty much the same thing but they are not. Have you ever tried tying your shoes really tight and as you are pulling on the laces, which increases tension, one of the laces snaps? It is pretty inconvenient when you have to replace a shoelace but think if that was your muscle! At times our muscles are capable of generating sufficient power to damage tendons or even break bones. They can cause avulsion, where the tendon tears off a piece of the bone at its attachment site. In order to prevent this, we have a safety mechanism in place called the Golgi tendon organ. Where we could consider the stretch reflex to be excitatory and cause contraction of the stretched muscle group the Golgi tendon reflex would be considered inhibitory and causes relaxation of the affected muscle. Therefore the result of activation of a GTO would be the opposite of the activation of a muscle spindle. The main purpose of GTOs is to prevent excessive tension on tendons and thus prevents injury.

Golgi tendon organs are composed of encapsulated nerve endings that are found interwoven with collagen fibers near the transition from muscle to tendon. These nerve endings monitor tension on the tendon rather than muscle length as muscle spindles do. As muscle contracts it develops tension on the tendon which is detected by the GTO. The GTO then sends action potentials, via afferent neurons, to the dorsal horn of the spinal cord where they synapse with inhibitory interneurons. The interneuron then synapses with and inhibits the alpha motor neurons in the anterior horn of the spinal cord. Inhibition of alpha motor neurons will effectively shut off the "power" to the muscle causing it to relax. You can think of this phenomenon almost like a circuit breaker. If there is a spike in power coming into your home that could potentially damage electrical devices the circuit breaker is tripped, temporarily shutting off electricity to those electrical devices.

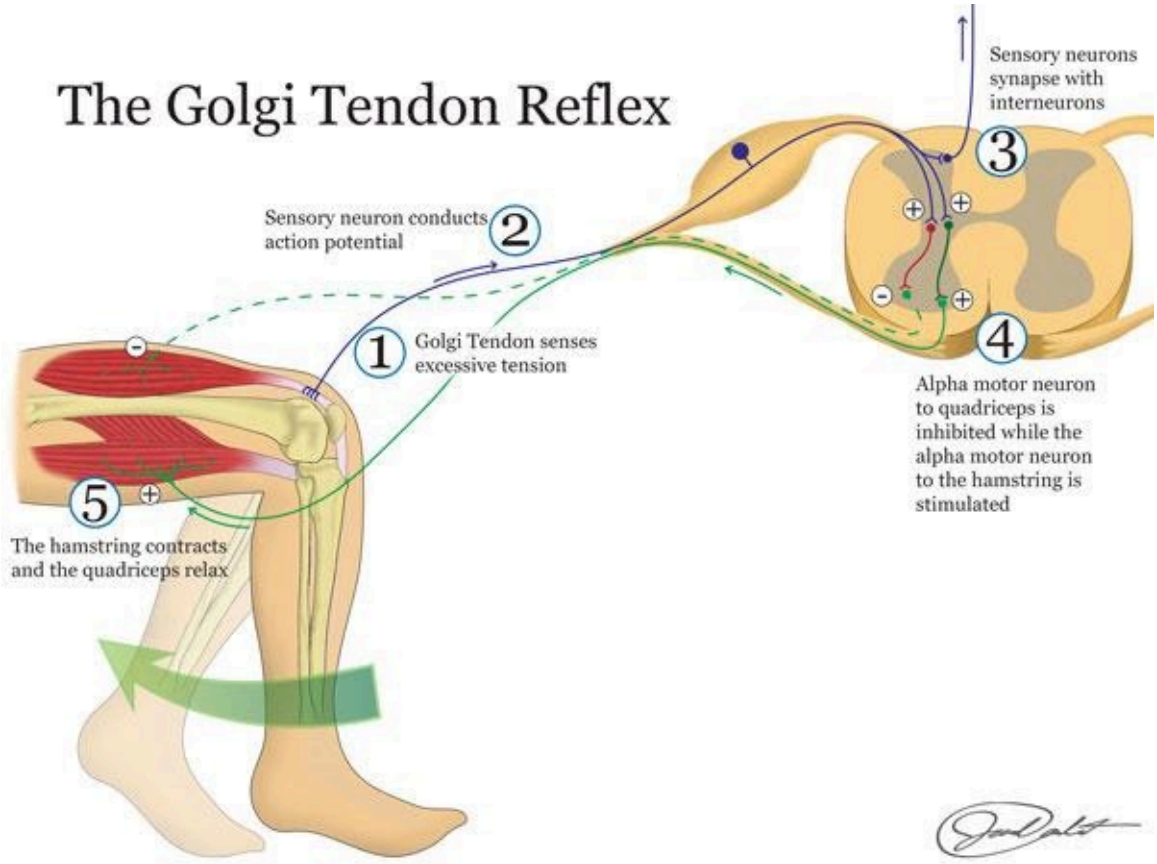


Image drawn by BYU-I student Jared Cardinet Winter 2015

You might ask yourself, "If this prevents excessive tension on muscles, what about those stories I have heard about mothers lifting cars off of babies and such?" Well, remember that this is a reflex and is generally managed from the bottom up without too much oversight from the upper motor neurons. In some circumstances, such as the super human feats of strength you have heard about, the CNS has the ability to override the reflex of the GTO. This happens as upper motor neurons modify the reflex at the level of the spinal cord. This allows extreme amounts of force and tension to be achieved, but the downside is that it usually causes pretty severe damage to the musculoskeletal system.

Withdrawal Reflex and Reciprocal Inhibition

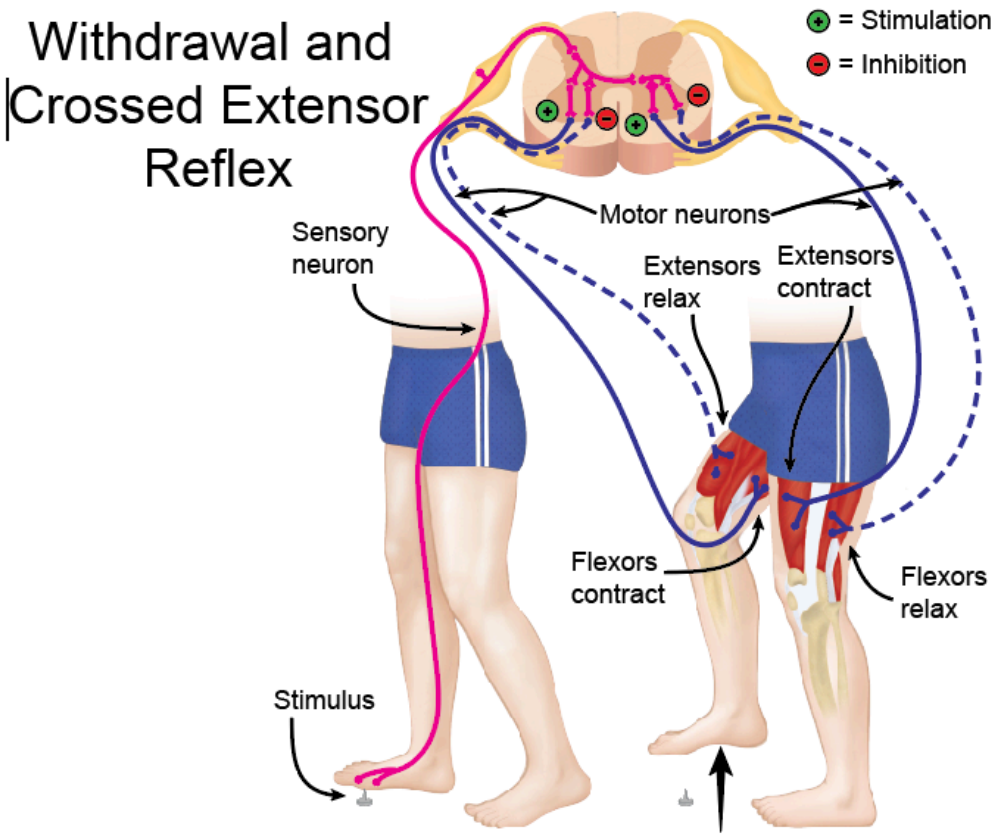


Image drawn by BYU-I student Nate Shoemaker Spring 2016

Have you ever stepped on something sharp with your bare feet or touched something hot with your hand? If the answer is yes then you have experienced the grace of the withdrawal reflex. If the answer is no, you need to live a little! The withdrawal reflex is yet another way that we are hard wired to avoid pain and tissue damage. We have free nerve endings, called **nociceptors**, scattered throughout our body that are sensitive to pain. When stimulated these sensory neurons activate lower motor neurons in the spinal cord. The lower motor neurons then stimulate contraction of skeletal muscle to remove or withdraw ourselves from the pain generator. In general, this will take place as flexor muscles are stimulated to contract, such as the hamstrings and hip flexors if you step on a tack or the biceps when you touch a hot stove. For this reason, the withdrawal reflex is sometimes called the flexor reflex.

In order for this to happen efficiently, we need to stimulate the flexor muscles and at the same time inhibit the extensor muscles. This phenomenon, called **reciprocal inhibition**, that was discussed in terms of the knee-jerk reflex is also at play here. The pain neuron, as it enters the dorsal horn of the spinal cord, will branch to stimulate an excitatory interneuron and an inhibitory interneuron. The excitatory interneuron then stimulates muscle contraction of the flexor muscle while the inhibitory interneuron causes the antagonist muscle, or the extensors, to relax.

Crossed Extensor Reflex

The crossed extensor reflex is yet another way that your body protects itself. When you step on that tack and reflexively pull your foot away you quickly find yourself supporting all of your weight on one leg. Without the crossed extensor reflex, instead of standing on one leg after stepping on a tack you would probably wind up on your backside.

Again, when you step on a tack and stimulate the pain fibers in your foot they send signals to the spinal cord through the dorsal horn. In addition to sending branches to excitatory and inhibitory interneurons on the same side of the body, the pain neuron also sends a branch to an excitatory interneuron that crosses over to the opposite side of the spinal

cord and stimulates a lower motor neuron. This lower motor neuron stimulates the extensor muscles on the opposite side of the body in preparation for the increased load as you shift your weight to that side.

SUMMARY

Now that you know what reflexes are and how they work, let's revisit the question, "How could the doctor tell that your friend was ok simply by looking in her eyes?" The answer is that he was checking the integrity of a different reflex, the pupillary light reflex. Remember, the clinical usefulness of checking reflexes is that specific stimuli should elicit predictable responses. Therefore, you would anticipate that shining a bright light in a person's eyes would cause the pupils to constrict, and that is exactly what should happen, but how? There are special receptors in the eye that are sensitive to light. When stimulated, like when the doctor shined the bright light in your roommate's eye, they transmit signals through the optic nerve to the midbrain. In the midbrain, these neurons stimulate the oculomotor nerves, which supply the muscles that cause constriction of the pupil. Thus by checking the pupillary light reflex, the physician was able to quickly evaluate the seriousness of the injury. In the case of severe brain injury, this reflex can be compromised so that the bright light would not cause the anticipated pupil constriction.

Some nice Youtube summaries of these reflexes are found at: https://books.byui.edu/-XAC

10.0

MODULE 10: THE AUTONOMIC NERVOUS SYSTEM

10.1

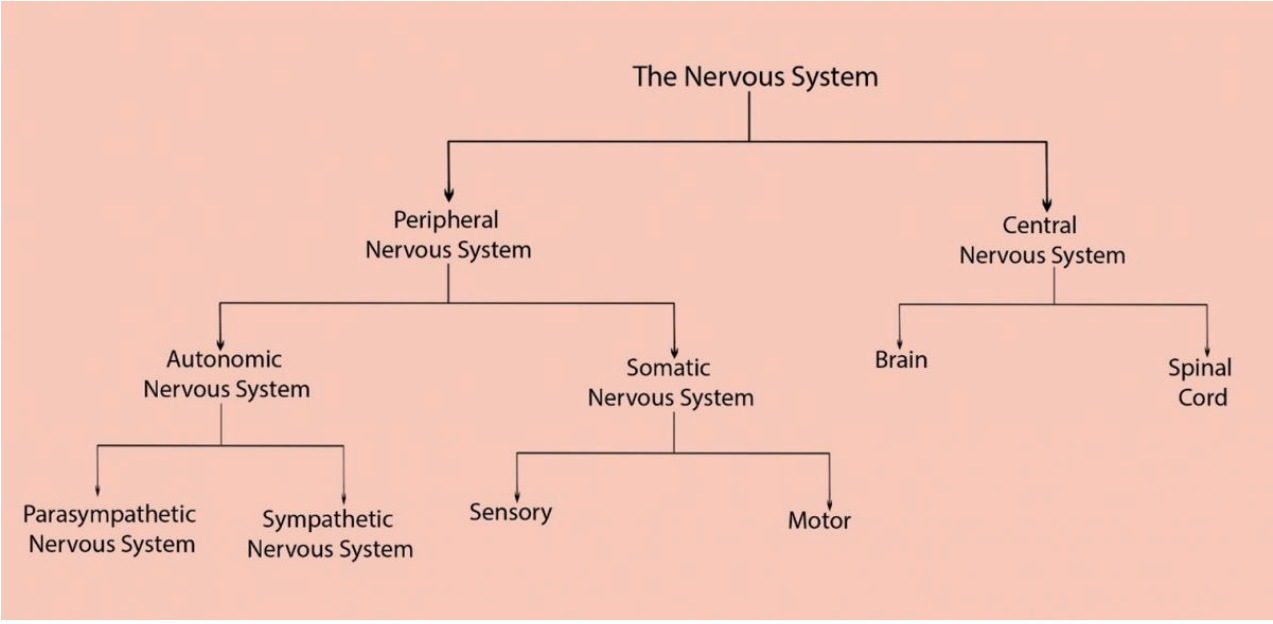
ORGANIZATION OF THE NERVOUS SYSTEM

10.1.1

Introduction to the Autonomic Nervous System

The weekend you've been anticipating for months is finally here-the big hiking trip with your pals in Grand Teton National Park. Thirty minutes into the hike you are confronted by a giant, grizzly mama bear and her cubs. Almost instantly your pupils dilate, your heart begins racing, the hair on the back of your neck stands up, and you feel a surge of energy as your adrenaline rises and you high tail it in the opposite direction. Your body's reactions to the bear can be summed up as a "fright, flight, or fight" reaction elicited by your central nervous system in response to incoming sensory information from the environment (i.e. seeing the bear, hearing the bear, and maybe even smelling the bear). During this response, your nervous system also slows down organs of digestion, urination, and defecation, so that all available energy may be used for running away. The actions on your heart, lungs, pupils, digestive system, and urinary system have occurred without your conscious awareness and are controlled involuntarily by the **autonomic nervous system** (**ANS**).

The human body is composed of one nervous system that can be subdivided into a **central nervous system** (CNS) and a **peripheral nervous system** (PNS) (see figure below). The brain and spinal cord make up the CNS, while the PNS is made up of any nervous tissue outside the brain and spinal cord, including 12 pairs of cranial nerves, 31 pairs of spinal nerves, and peripheral sensory receptors. The PNS can be further subdivided into the ANS and the somatic nervous system depending on which type of muscle it innervates and whether or not it is voluntarily controlled.



Organization of the Nervous System. *Image by BYU-Idaho student, 2013*

The ANS can be subdivided into **sympathetic** and **parasympathetic nervous systems**. The ANS neurons innervate smooth muscle, cardiac muscle and glands. The ANS efferent neurons do not innervate skeletal muscle. It is the sympathetic branch of the ANS that is responsible for the "fright, flight, or fight" response elicited by the encounter with

the bear. Another example of a sympathetic response that might be a little closer to home would be your body's reaction after being jilted by your fiancé at the altar. You can imagine that your heart rate would increase, you might start to hyperventilate, and you definitely would have no desire to eat. A sympathetic response can also occur during illness or physical trauma, from anxiety, or pretty much any stressful situation. Such a response is characterized by increased heart rate and blood pressure, goosebumps, pupil (pupil dilation=**mydriasis**), bronchiole dilation, and increased blood flow to cardiac and skeletal muscles.

The parasympathetic division, on the other hand, is responsible for energy conserving **("rest and digest")** activities, including decreased heart rate, blood pressure, and respiration; constriction of the pupil (**miosis**); increased secretions and peristalsis of the digestive tract; and increased urination. The acronym **SLUD (S**alivation, **L**acrimation, **U**rination, and **D**efecation) may be useful to remember some of the responses caused by the parasympathetic division in certain organs. Other than some sweat glands, most secretions of the body increase when the parasympathetic nervous system is activated.

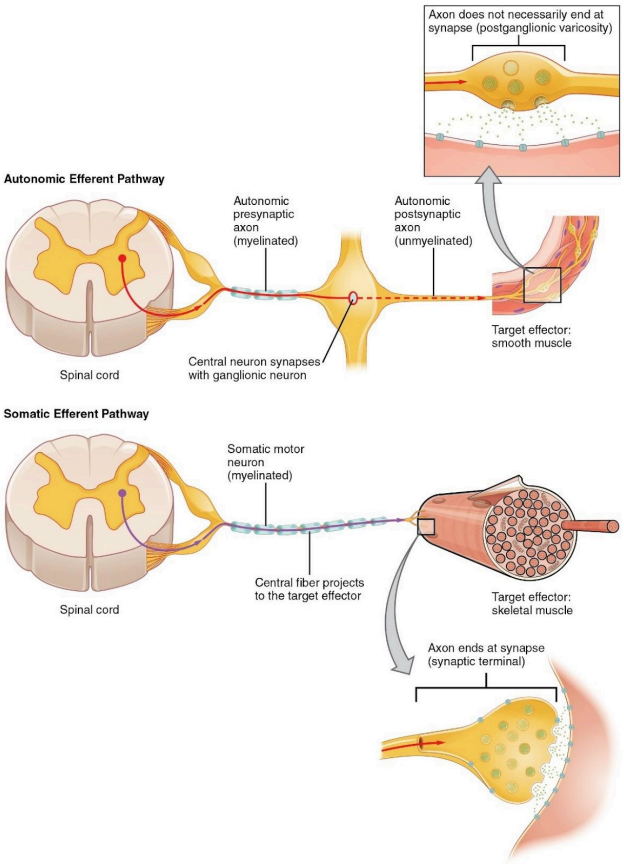
The ANS innervates visceral organs - organs which are unconsciously controlled by the brain. Visceral organs contain either smooth or cardiac muscle; respective examples include the intestines and the heart. Interestingly, if a visceral organ is removed from the body and placed in an oxygenated Ringer's solution, it will continue to undergo peristalsis (wave-like smooth muscle contractions of the gastrointestinal tract) or beat without even being connected to the ANS. This is called auto rhythmicity. Why then, do you ask, is the ANS even necessary for these organs to function? The answer is, it is not. But, the ANS is necessary to regulate the activity of these organs, essentially causing them to speed up or slow down in order to maintain homeostatic conditions in the body.

Usually, each visceral organ is innervated by nerves from both sympathetic and parasympathetic divisions, and effects of these divisions are most often in opposition to one another. This type of "wiring" is called **dual autonomic innervation**. The heart is a good example of this. It is innervated by fibers from both parasympathetic and sympathetic divisions that oppose one another. Increasing parasympathetic stimulation to the heart will cause decreased heart rate while increasing sympathetic activity will increase heart rate and force of contraction.

10.1.2

Structural Organization and Anatomy of the ANS

The ANS uses a “two neuron system” to relay electrical signals from the CNS to effectors (organs, glands, and vessels). This is different from the somatic motor division where just one neuron extends from the CNS to skeletal muscle. Parasympathetic and sympathetic divisions are “wired” similarly in that they both have a preganglionic neuron and postganglionic neuron. (Recall from a previous module that a ganglia is a collection of cell bodies located outside the CNS). These neurons get their names from their anatomical location in relation to autonomic ganglia, or relay centers.

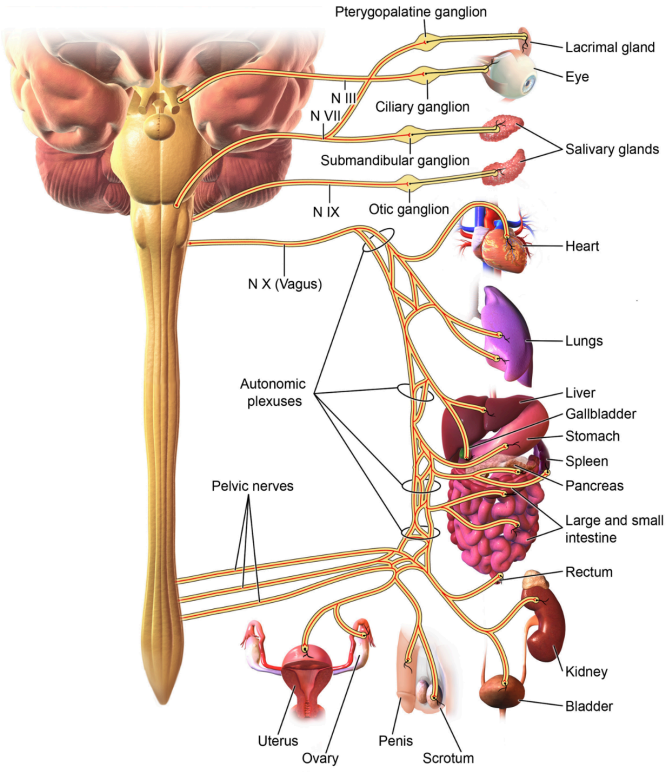


Comparison of Somatic and Autonomic Pathways. *By OpenStax College [CC BY 3.0 (http://creativecommons.org/licenses/by/3.0)], via Wikimedia Commons*

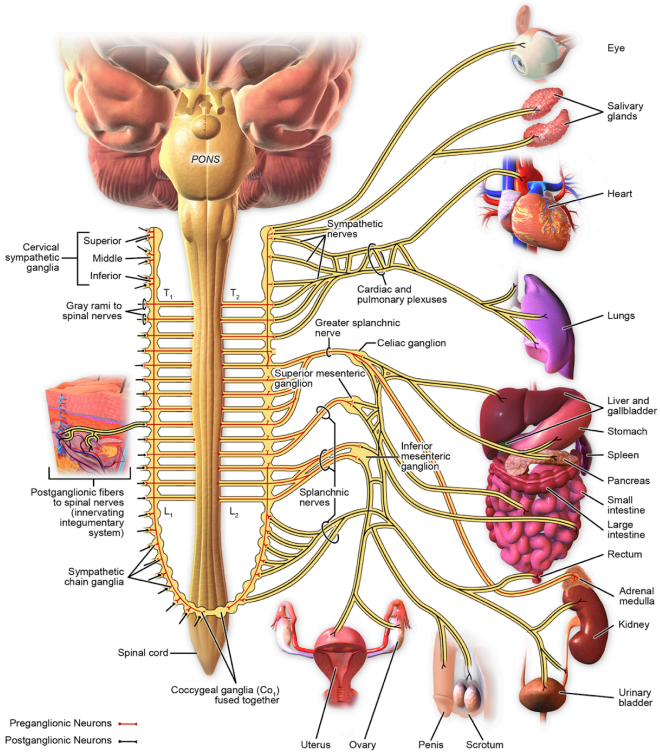
Autonomic ganglia for the sympathetic division include – **sympathetic chain ganglia** which are located near the spinal column (labeled at the bottom of the chain in the SNS anatomy image below) and **collateral ganglia**. The collateral ganglia in the image below are located further away from the spinal column and are labeled Celiac, Superior mesenteric and Inferior mesenteric. Autonomic ganglia for the parasympathetic division are called **terminal ganglia** and these are

located very near the effector they innervate. They are not labeled as there are so many places that parasympathetic neurons synapse with post ganglionic neurons in the very walls of the organs themselves.

The cell bodies for preganglionic neurons are located in either the brain stem or spinal cord, and their axon terminals are located in autonomic ganglia. In the ganglia, a neuron-to-neuron synapse relays information to the cell body of a postganglionic neuron. Postganglionic neurons, also located in the autonomic ganglia, then transmit the signal to effectors. The synapse between the postganglionic neuron and the effector is known as a neuroeffector synapse or neuroeffector junction.



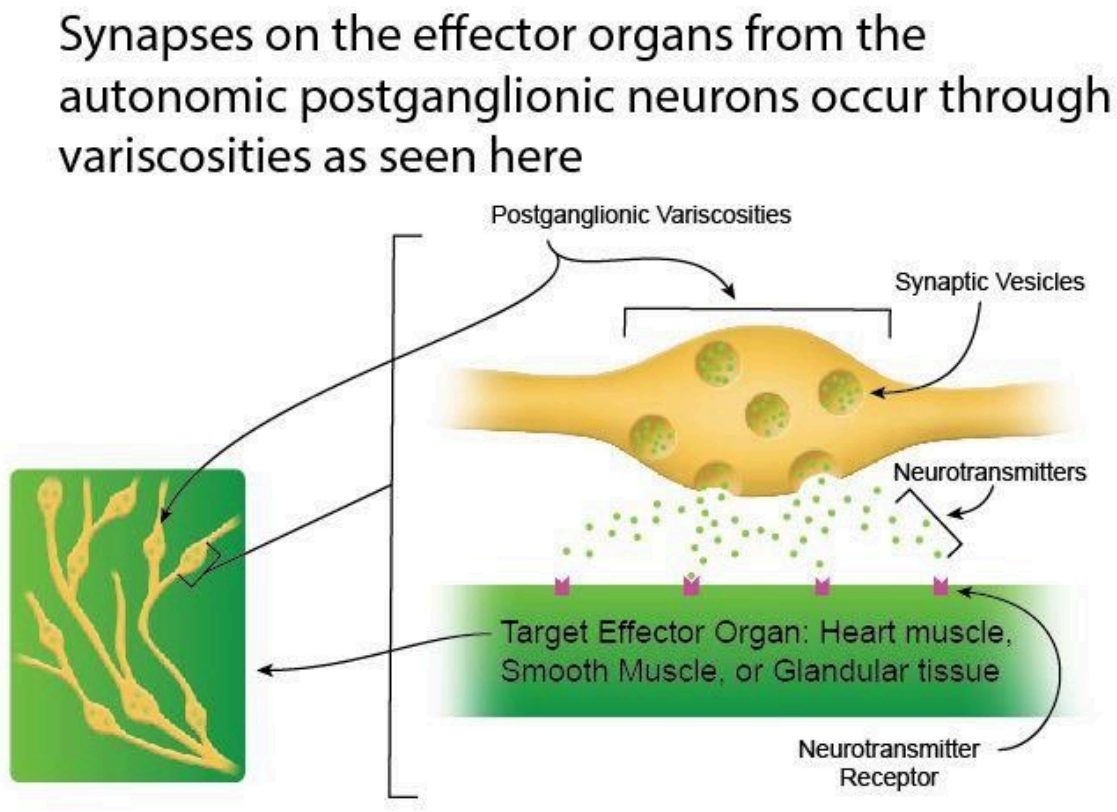
Parasympathetic Innervation



Sympathetic Innervation

Downloaded from Wikimedia Commons Dec 2013; Author: BruceBlaus; License: Creative Commons Attribution 3.0 Unported license.

In general, the sympathetic division uses shorter preganglionic neurons and longer postganglionic neurons while the parasympathetic division uses long preganglionic neurons and short postganglionic neurons. Postganglionic release neurotransmitter onto effector organs. However, the synapse is a little unique. Rather than forming a nerve terminal at a synaptic junction, we see many swellings (or varicosities) develop on the most distal segments of the postganglionic neuron and they secrete neurotransmitter onto the effector tissue (see image below)



Autonomic Postganglionic Neuron Synapse. *Image by BYU-Idaho student Nate Shoemaker Spring 2016*

10.1.3

The SNS and the PNS

The table below helps us compare and contrast some of the characteristics of the SNS and the PNS.

	Sympathetic		Parasympathetic	
	Preganglionic Neuron	**Postganglionic Neuron**	**Preganglionic Neuron**	**Postganglionic Neuron**
Neuron Length	Short to Medium	Medium to Long	Long	Short
Neurotransmitter Released	ACH	NE (except sweat glands and some blood vessels – ACH)	ACH	ACH

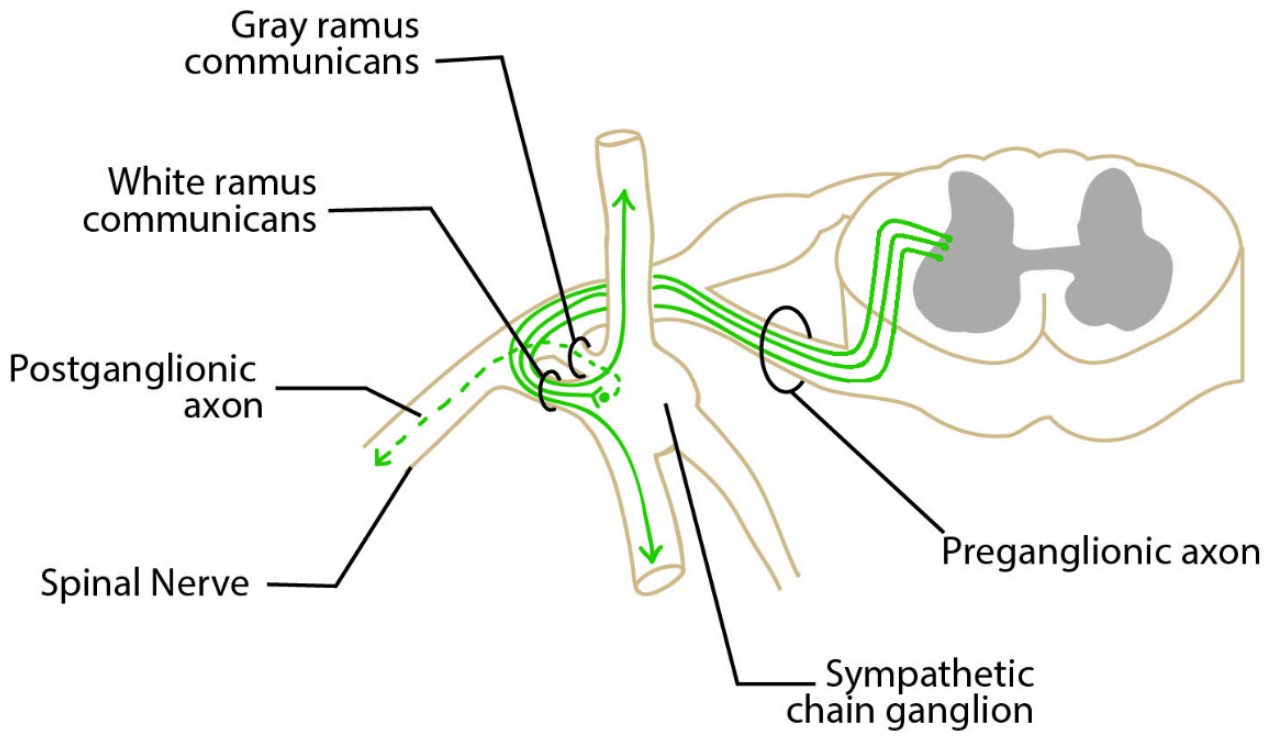
Characteristics of Sympathetic and Parasympathetic Nervous System. *Image by BYU-Idaho Student 2013*

ACH is short for Acetylcholine and NE is short for Norepinephrine. Acetylcholine and Norepinephrine are neurotransmitters.

Sympathetic Division (SNS)

The cell bodies of the preganglionic axons of the sympathetic division are located in segments T1 through about L2 to L3 of the lateral horn of the spinal cord. From here, these axons project away from the spinal cord through the ventral root and enter a spinal nerve. They then exit the spinal nerve through a white ramus communicans (myelinated axons) and enter a sympathetic chain ganglia, which are ganglia located along the spinal cord bilaterally. The following are descriptions of four different routes taken by sympathetic axons traveling from the CNS, to their **effectors** (organs, glands, and vessels) (see figures below).

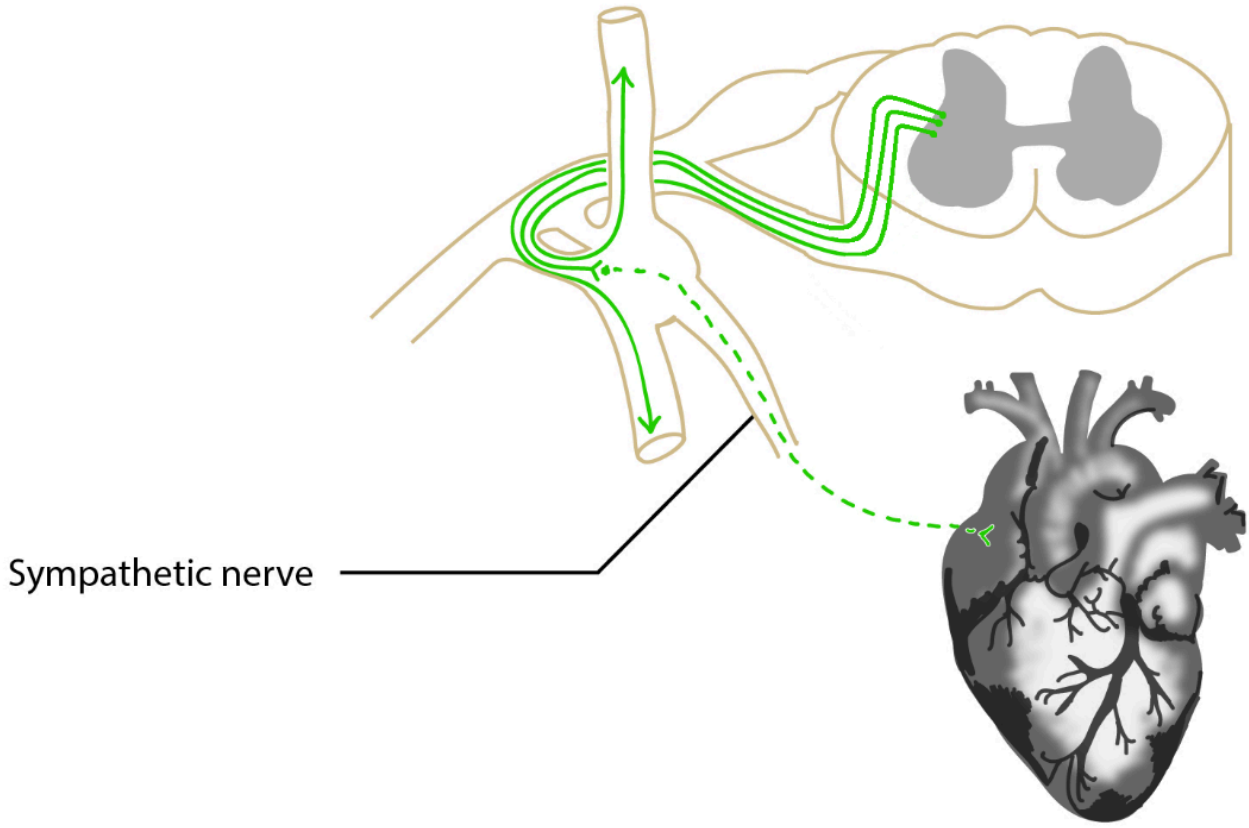
1. Preganglionic axons synapse at the sympathetic chain ganglia with a postganglionic neuron. The postganglionic neuron then leaves the sympathetic chain ganglia through a gray ramus communicans (unmyelinated axons) and reenters the **spinal nerve** and travels to the skin and blood vessels throughout the body.



Sympathetic Chain Ganglion. *Image by BYU-Idaho student, Kaylyn Lloyd Winter 2014*

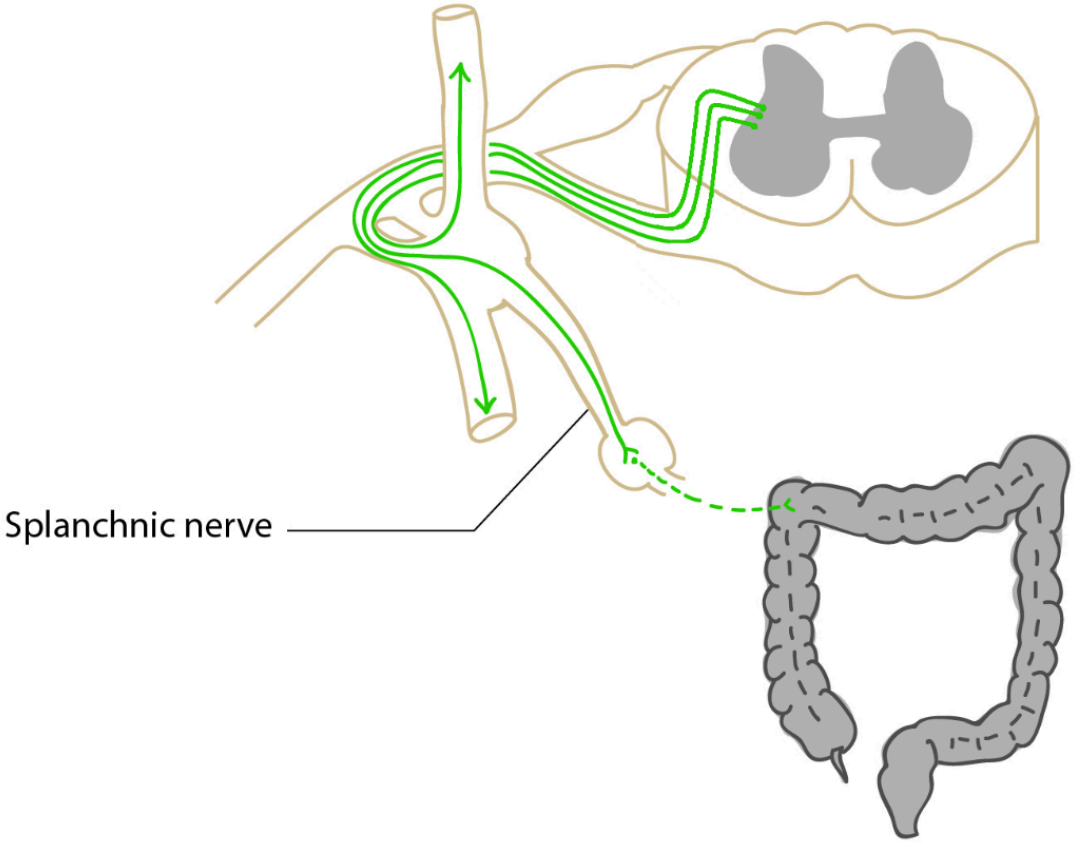
In the image above, preganglionic axons enter a sympathetic chain ganglion via a white ramus communicans (called white because the axons are myelinated which gives a more whitish appearance to this "bridge). Some axons synapse with postganglionic neurons in the sympathetic chain ganglion, while others travel to inferior or superior sympathetic chain ganglia before synapsing. Postganglionic axons leave sympathetic chain ganglia via a gray ramus communicans (called gray because the postganglionic neurons are not myelinated which gives a grayish appearance to the "bridge") and enter a spinal nerve. The spinal nerve carries the postganglionic axon out the peripheral body with other sensory and motor neurons.

2. The second type is very similar, but instead of the postganglionic neuron entering a spinal nerve, it enters a **sympathetic nerve** and travels to organs of the thoracic cavity. See below.



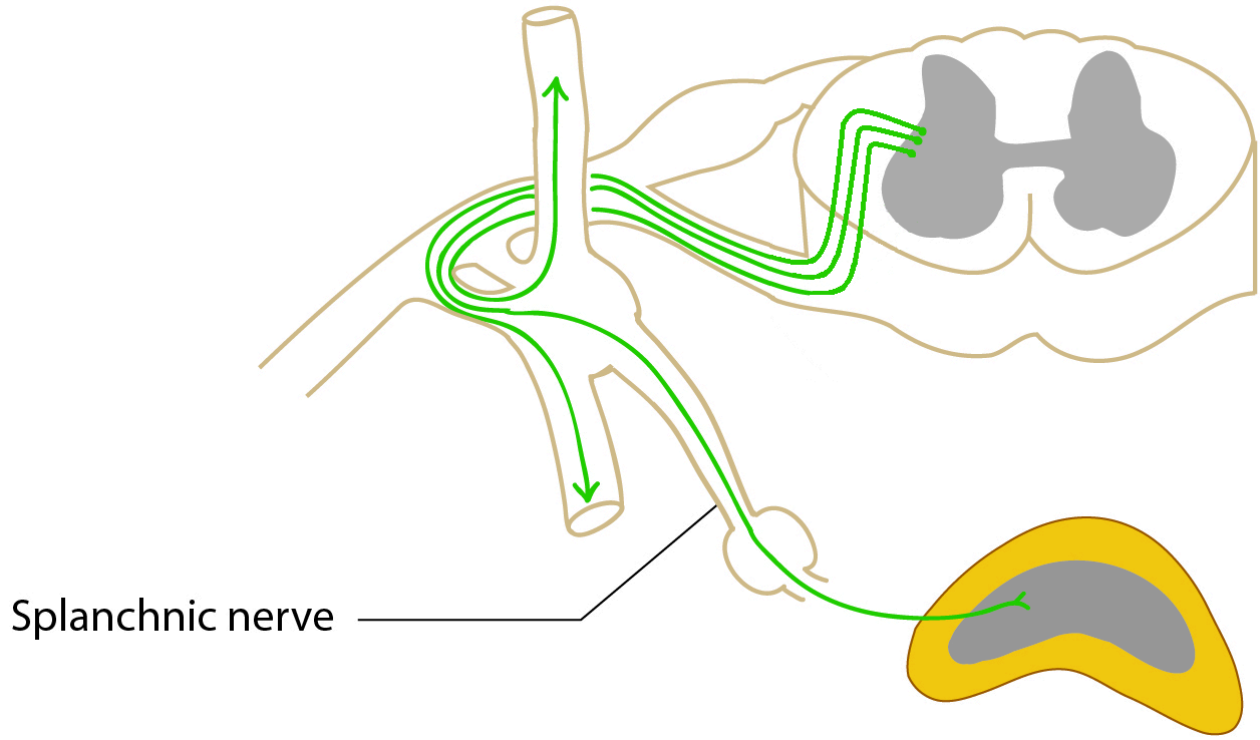
Sympathetic Nerve Signaling Heart. *Image by BYU-Idaho student, Kaylyn Lloyd Winter 2014*

3. The preganglionic neuron enters and leaves the sympathetic chain ganglion without synapsing and forms a **splanchnic nerve** and travels to collateral ganglia. At these ganglia, the preganglionic neurons synapse with postganglionic neurons which then extend to organs, glands, and vessels of the abdominopelvic cavity. See below.



Splanchnic Nerve Acting on Intestines. *Image by BYU-I student, Kaylyn Lloyd Winter 2014*

4. The last route for sympathetic axons is similar to those traveling through splanchnic nerves, but instead of synapsing, they travel straight through collateral ganglia. They then extend to the **medulla of the adrenal gland**, where they synapse with cells that produce epinephrine (EPI) and norepinephrine (NE). These medullary cells function as modified postganglionic neurons and release secretory product directly into the blood rather than into a synapse. About 80% of adrenal medullary cells produce EPI and the other 20% produce NE. After release into the blood, these hormones travel to receptors throughout the body to elicit a "fright, flight, or fright" response. See below.



Splanchnic Nerve Signaling Medulla of the Adrenal Gland Releasing Neurotransmitters directly into the Blood Stream for a System-wide Sympathetic Response. *Image by BYU-I student, Kaylyn Lloyd Winter 2014*

Click on this highlighted link to follow an image search that will show some more pictures for pathways of the sympathetic nervous system.

About 8% of the fibers in the 31 pairs of spinal nerves are postganglionic sympathetic fibers. Some of these fibers innervate the effectors of the skin-particularly capillaries and sweat glands. The sympathetic division also innervates the iris lens, nasal mucous membranes, salivary glands, the heart, lungs, stomach, intestines, adrenal gland, and urinary bladder (see figure below).

Parasympathetic Division (PNS)

The parasympathetic division does not follow 4 pathways like the sympathetic division. The parasympathetic division sends preganglionic neurons from the cranial area and the sacral area. This is why it is also known as the craniosacral division.

Cranio: Preganglionic cell bodies for coming from the brain are located in the brainstem and make up part of the cell bodies of cranial nerves - namely, cranial nerves III (oculomotor) which control the size of the pupil and shape of the lens, VII (facial) which control nasal mucous membranes and lacrimal and salivary glands, IX (glossopharyngeal) which controls the parotid salivary gland, and the X (vagus) which innervates organs of the thoracic cavity and upper abdominal cavity including the lungs, heart, stomach, pancreas, small intestine, liver, and upper portion of the large intestine. The vagus nerve is the major nerve of the cranial parasympathetic division. 75-80% of all parasympathetic fibers are found in the vagus nerve.

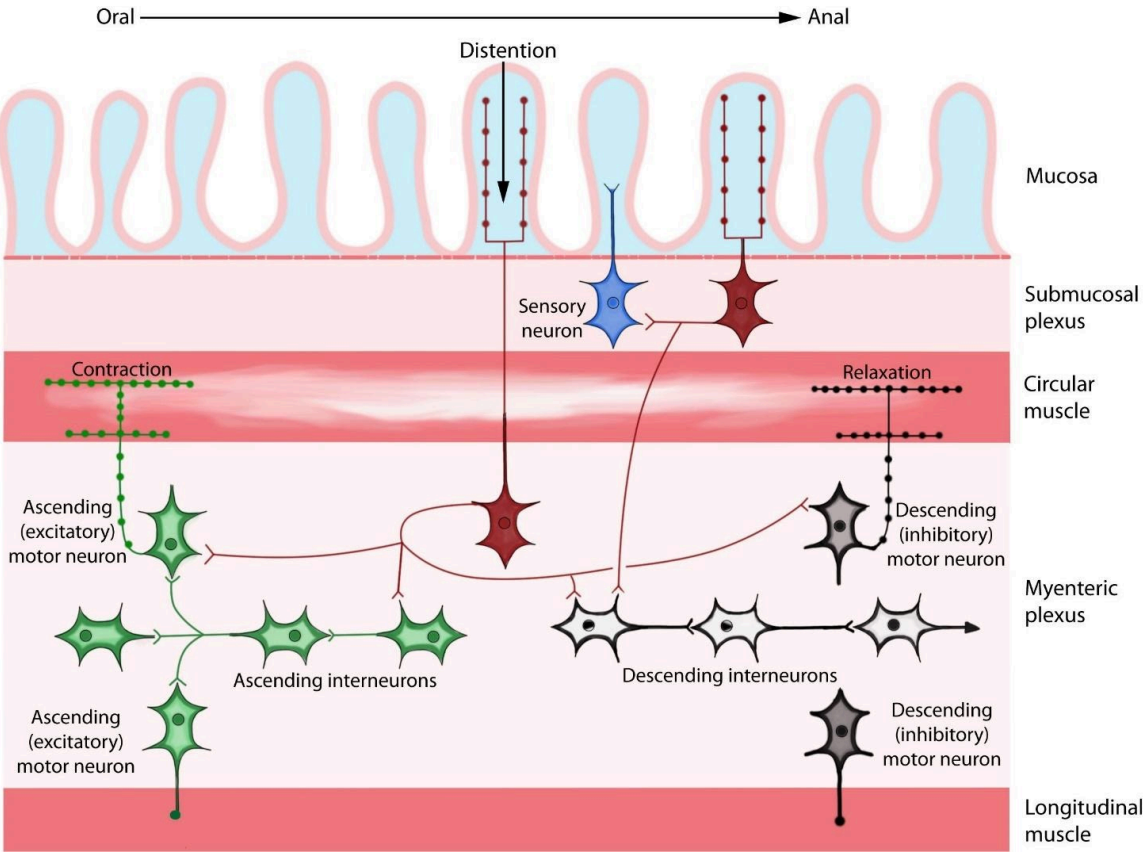
Sacral: There are a few preganglionic neuron cell bodies of the parasympathetic division that are located in the sacral region of the lateral horn of the spinal cord. Axons from these neurons enter **pelvic splanchnic nerves** and then extend to **terminal ganglia** which are located near or on the effector. Effectors innervated by the lower portion of the parasympathetic nervous system include the lower half of the large intestine and organs of the reproductive and renal systems. It helps to have images of the ANS anatomy in hand as you study these ANS divisions.

It would be a good idea to review the ANS anatomy pictures presented in 10.1.2.

10.1.4

The Enteric Nervous System

The enteric nervous system (ENS) is sometimes referred to as the third division of the nervous system (central, peripheral, and enteric). This system is composed of a nerve plexus or a meshwork of fibers innervating the digestive tract from the esophagus to the distal colon. The ENS includes the **myenteric plexus** and the **submucosal plexus** which receive preganglionic fibers from the parasympathetic division and postganglionic fibers from the sympathetic division of the ANS. Innervation from the ANS and sensory input from within the wall of the gut work together to control smooth muscle motor activity and gut secretory actions. However, the ENS releases a variety of neurotransmitters and is capable of controlling digestive functions independently of the CNS by way of local reflexes. When food is introduced into the digestive tract, stretch receptors in the gut are activated and send action potentials through afferent **enteric sensory neurons**. These neurons synapse with **enteric interneurons** which are capable of activating efferent **enteric motor neurons**. These neurons innervate glands and smooth muscle. Their increased activity enhances digestive enzyme secretions and gut contraction to cause mixing and propulsion of food. The ENS is particularly important in providing synchronous peristaltic movements ensuring propulsion of food in one direction (see figure below).



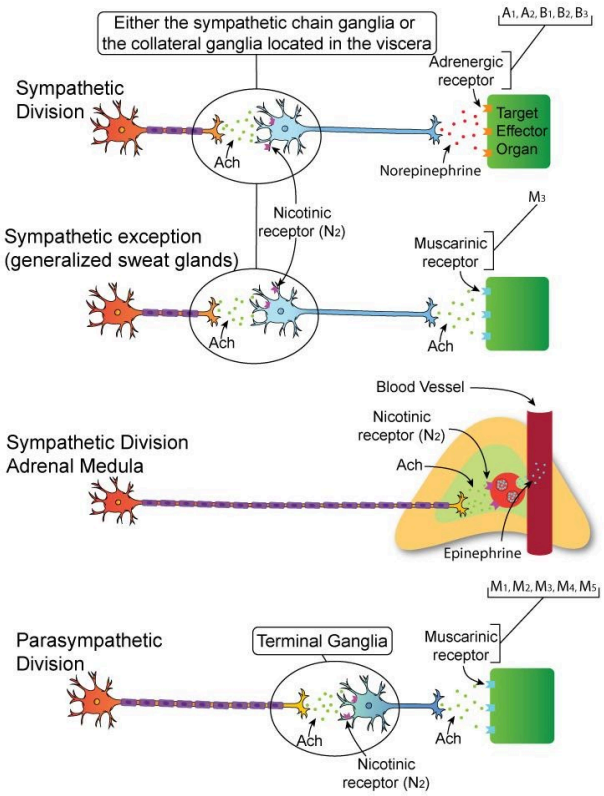
The enteric nervous system, showing the submucosal and myenteric plexuses.

Image by BYU-Idaho student 2014

10.2

PHYSIOLOGY OF THE ANS

As discussed previously, nerves of the ANS extend from the CNS to smooth, or cardiac muscle, organs, and glands via a two neuron system—namely a preganglionic neuron and a postganglionic neuron. In this system, there are 2 synapses—one separating preganglionic and postganglionic neurons, and the other between the postganglionic neuron and effector (see figure below). The preganglionic neuron releases neurotransmitters stored in synaptic vesicles of axon terminals. An action potential reaching the axon terminal causes the release of these stored neurotransmitters into the synaptic cleft. After crossing the synapse, neurotransmitters bind to receptors imbedded in postganglionic cell membranes (see figure below). This binding depolarizes the postsynaptic cell membrane (EPSP) and results in action potentials. Action potentials arriving at the axon terminal lead to the release of neurotransmitters into the synaptic cleft, separating the postganglionic neuron and effector (neuroeffector junction). Finally, binding of the neurotransmitter to the receptor expressed on effector cells can result in excitation or inhibition of the effector. Sympathetic and parasympathetic divisions differ in the types of neurotransmitters they release and the receptors and second messenger systems they express.



ANS preganglionic and postganglionic neurons and locations of ANS receptors. *Image by Nate Shoemaker Spring 2016*

Nicotinic receptors are located on the postganglionic neurons of the sympathetic and parasympathetic cell bodies. Nicotinic receptors respond to the binding of acetylcholine (ACH), which causes an excitatory effect. Muscarinic receptors are located on all parasympathetic effector cells and some (generalized sweat glands) sympathetic effector cells. Muscarinic receptors respond to the binding of ACH, and may have an excitatory or inhibitory effect. Adrenergic receptors are located on most sympathetic effector cells. Adrenergic receptors respond to the binding of norepinephrine (NE), which may have an excitatory or inhibitory effect.

Neurotransmitters of the ANS

Receptors of the ANS

10.2.1

Neurotransmitters of the ANS

Neurotransmitters are chemicals that travel across the synapse connecting two neurons, or between a neuron and an effector. For example, when we discussed the neuromuscular junction we were talking about a neuron-effector synapse and the neurotransmitter used was **acetylcholine (ACH)**. ACH is also one of the neurotransmitters used by the ANS. **Cholinergic neurons** produce ACH and store ACH in their synaptic terminals. The preganglionic neuron for both parasympathetic and sympathetic nervous systems is cholinergic. The postganglionic neuron of the parasympathetic division is also cholinergic. The postganglionic neuron for the sympathetic division is usually an **adrenergic neuron** which means that it produces catecholamines. Catecholamines are an organic chemistry group that includes **norepinephrine (NE), epinephrine (EPI) and dopamine**. In the Sympathetic nervous system NE is the neurotransmitter found at the synapse between postganglionic neurons and the organ. Sympathetic postganglionic neurons innervating sweat glands and some reproductive system blood vessels are the exception; they are cholinergic and release ACH.

10.2.2

Receptors of the ANS

Usually, a particular receptor subtype for each division of the ANS will dominate in a certain gland or organ. In general, activation of some receptor subtypes leads to stimulation of the effector and activation of others to inhibition of the effector. Even numbered subtypes are usually inhibitory and odd numbered subtypes are usually excitatory, but there are no hard and fast rules. Ultimately, the relative amounts of each receptor subtype expressed in the tissue will determine the overall effect (stimulation or inhibition) on the particular gland or organ.

Cholinergic Receptors

As mentioned, preganglionic neurons of both sympathetic and parasympathetic divisions produce and release ACH. The receptors for ACH are known as cholinergic receptors. There are two main subtypes of cholinergic receptors-; nicotinic and muscarinic. They are named after alkaloids found in tobacco and certain mushrooms respectively. The alkaloid nicotine specifically activates nicotinic cholinergic receptors, while muscarine activates muscarinic cholinergic receptors, and ACH activates both types. The cell bodies of postganglionic neurons for both sympathetic and parasympathetic nervous systems express nicotinic receptors (see figure above). To distinguish nicotinic receptors in neurons from nicotinic receptors found in the neuromuscular junction, we use the terms **nicotinic (N1 or N2)** cholinergic receptors. N1 are located in the neuromuscular junction and N2 are used in the ANS. Similar to the neuromuscular junction, stimulation of nicotinic type II (N2) channels results in the entry of Na^+ which depolarizes the post synaptic neuron.

Muscarinic receptors (M) are located on cells of all parasympathetic effectors and on cells of some sweat glands innervated by the sympathetic nervous system. There are several subtypes of muscarinic receptors **(M1-M5)** which may be stimulatory (depolarization) or inhibitory (hyperpolarization)

Adrenergic Receptors

As mentioned, neurons that produce and release the neurotransmitter NE are known as adrenergic neurons. NE is secreted by postganglionic neurons of the sympathetic nervous system and binds to **adrenergic receptors** expressed on effector cells. Epinephrine (EPI) released by the adrenal gland also binds to adrenergic receptors expressed on effectors (see figure above). There are two main types of adrenergic receptors, namely, alpha and beta which have several subtypes. For our purposes, we will focus on the following five subtypes: alpha **1, 2, and beta 1, 2, and 3**. Activation of adrenergic receptors expressed on effectors by NE or EPI may result in stimulation or inhibition of the effector depending on the tissue involved. Odd subtypes of adrenergic receptors (alpha 1, and beta 1, and 3) generally have stimulatory effects and even subtypes (alpha 2 and beta 2) have inhibitory effects. NE has a stronger affinity for alpha 1 rececptors than EPI and EPI has a stronger affinity for Beta-2 receptors than NE. The next section explores the effects of a ligand binding to its receptor in a particular tissue.

10.3

ACTIONS OF THE AUTONOMIC NERVOUS SYSTEM

Certain drugs exert their effects by binding to cholinergic and adrenergic receptors to increase or decrease the activity of effectors normally controlled by the ANS. Drugs that are **agonists**, binds to a specific receptor and activates it, while an **antagonist** binds to a receptor and prevents it from being activated, or inhibits it.

Another drug mechanism that can affect the autonomic nervous system involves inhibition of enzymes that break down the normally secreted neurotransmitters. For example, acetylcholinesterase inhibitors cause an excessive amount of acetylcholine to build up in the synaptic cleft (because it is not broken down by the enzyme called acetylcholinesterase). This excessive acetylcholine causes more parasympathetic symptoms than sympathetic symptoms because acetylcholine is used in both the first and second synapses of the two neuron pathway to body target tissues. Acetylcholinesterase inhibitors do affect the sympathetic nervous system in the first synapse but not the second synapse where norepinephrine is released. Therefore if a person is exposed to an acetylcholinesterase inhibitor they will have symptoms like bradycardia, diarrhea, bronchoconstriction and constriction of the pupils. Interestingly, they will have a lot of general body sweating as well even though that is a sympathetic response. But, that is because general sweating is the one exception in the sympathetic nervous system that has acetylcholine in the second synapse between neuron and target organ. Sweating is the one exception that acts just like the parasympathetic nervous system in its organization of the neurotransmitters and target receptors.

In the following section, some other drugs will be presented. Having an understanding of the ANS and its particular receptors located on effectors and the drugs that activate or block these receptors will assist your understanding of the actions of this system.

Eyes

The eye has multiple autonomic functions controlled by several autonomic receptors. Among these are the intrinsic muscles of the eye (those controlling the size of the pupil and the shape of the lens) and the secretory epithelium (produces aqueous humor) of the ciliary body.

Circular and radial muscles of the iris, named sphincter pupillae and dilator pupillae respectively, control how much light enters the eye. Outer iris smooth muscles - the dilator pupillae muscles express alpha 1 receptors, cause mydriasis when they contract and are controlled by sympathetic fibers. The inner, sphincter pupillae muscles are innervated by the parasympathetic division, express M3 receptors, and cause miosis when they contract.

Ophthalmologists often need to enlarge the diameter of the pupil in order to more easily examine the retina. **Phenylephrine** is an alpha 1 agonist and **atropine** is a muscarinic antagonist. Both are mydriatics and are administered as eye drops to reduce systemic effects.

The function of the lens is to focus an image on the retina. Depending on whether light rays are coming from an object seen up close (more bending of light required) or far away (less bending required), the lens changes shape to allow for this clear focus. The lens of the eye is an elastic bi-convex structure made of crystalline protein. The lens in a more spherical shape (more convex) will cause more bending of light which is necessary to see things close up. The lens in a

more flattened state (less convex) will cause less bending of light. A flatter lens is necessary to see things far away. Regulation of the ciliary muscles helps determine the convexity of the lens. Ciliary muscles are innervated by both parasympathetic and sympathetic fibers. Activation of beta 2 receptors expressed on ciliary muscles causes a flatter lens for far vision, while muscarinic receptors mediate a more convex lens for near vision. One of the side effects of atropine eye drops used to cause mydriasis is blurred vision because it impairs the lens' ability to accommodate for near vision (**cycloplegia**).

Contraction of the ciliary muscle also puts tension on the trabecular network. This action opens up its pores and facilitates outflow of aqueous humor into the canal of Schlemm and back into systemic circulation. For this reason, eye drops that are muscarinic agonists, such as **pilocarpine**, can be used to treat elevated intraocular pressure (glaucoma), lowering intraocular pressure by increasing the outflow of aqueous humor. Stimulation of beta 1 receptors on the ciliary body epithelium increases the production of aqueous humor. Therefore, beta 1 antagonists such as **betaxolol** are also often used to treat glaucoma, since they reduce the production of aqueous humor.

Blood Vessels

Arterioles of the body mostly express alpha 1 receptors on their smooth muscle cells. Activating these receptors results in an increase in intracellular calcium causing smooth muscle contraction. This contraction narrows the diameter of the arteriole lumen thus reducing blood flow. Since arterioles express primarily alpha 1 receptors, you might imagine that an increase in sympathetic nerve firing would result in vasoconstriction of most arterioles. Arterioles of certain organs including skeletal muscle and cardiac muscle express beta 2 receptors in addition to alpha 1 receptors. Beta 2 receptors are activated primarily by circulating epinephrine and their stimulation causes relaxation of smooth muscle and vasodilation. The degree of vasodilation is dependent on the density of alpha 1 vs. beta 2 receptors expressed on arterioles in a particular tissue as well as on the concentration of epinephrine in the blood. During a "fright, flight, or fight" response, vasodilation of certain arterioles supplying skeletal tissue is essential. Ample oxygen and nutrients are critical when running away from ferocious bears or scary dating situations.

Since alpha 1 receptors are so important in regulating the size of arterioles, activating or blocking them can greatly influence blood pressure. **Prazosin** is an alpha 1 antagonist used to treat high blood pressure. It can also be used to treat Raynaud's disease which results from excessive vasoconstriction particularly in the fingers, cutting off the blood supply leading to cold fingers and in severe cases gangrene.

While it is true that there are some blood vessels (very few) that have parasympathetic cholinergic innervation that will cause vasodilation, the vast majority of blood vessels have no parasympathetic innervation.

Sweat Glands

Sweat glands are exclusively innervated by the sympathetic division. Postganglionic neurons of the sympathetic division that innervate glands responsible for generalized sweating secrete ACH. This is the exception to the rule since postganglionic sympathetic neurons usually secrete norepinephrine. After its release from the postganglionic cell, ACH crosses the neuroeffector junction and binds to muscarinic receptors expressed on sweat glands for generalized sweating. Localized sweat glands are activated by stress and are those located in the palms, soles, genitalia, and armpits and express alpha 1 receptors. **Terazosin** is another alpha 1 antagonist that is sometimes used to treat excessive sweating (hyperhidrosis).

Heart

Activation of the sympathetic division and release of catecholamines from the adrenal medulla leads to increased heart rate and force of contraction. This stimulatory effect is due to a high concentration of beta 1 receptors in the myocardium and cells in the SA node. Selective beta 1 antagonists like **atenolol** are often used to treat high blood pressure by decreasing heart rate and force of contraction.

Parasympathetic stimulation or administration of a muscarinic agonist has an opposing effect on the heart, decreasing heart rate. Injectable **atropine**, a muscarinic antagonist, is often used with other drugs in emergency medicine to start

the heart back up after cardiac arrest. This action blocks parasympathetic activity which normally slows the heart rate.

Lungs

Activation of muscarinic receptors located in the smooth muscle lining the bronchiole tree results in constriction of air passageways, while activation of beta 2 receptors by circulating epinephrine causes smooth muscle relaxation and dilation of the bronchioles (Note: beta 2 receptors are not innervated by postganglionic fibers and therefore respond to circulating epinephrine secreted by the adrenal medulla). Pharmacological treatment aimed at opening up the airways focuses on blocking parasympathetic actions or augmenting actions of the sympathetic division. A muscarinic antagonist such as **ipratropium**, or a beta 2 agonist like **albuterol** can be administered via an inhaler and cause bronchodilation. This relaxation of the smooth muscle in air passageways is very important in the treatment of asthma and chronic obstructive pulmonary disease (COPD) when ventilation of the lungs is compromised.

Stomach and Intestines

Activity of the enteric nervous system can be modified by activity of the ANS. The gastrointestinal tract is dually innervated by both divisions, but regulation is not equal. Recall that the parasympathetic division is most active under "rest and digest" conditions. Parasympathetic fibers leading to the gastrointestinal tract are much more extensive and have a much greater influence on digestion compared to the sympathetic division. The parasympathetic division increases the secretions from glands, promotes mixing of food with digestive enzymes and bile, and propels material down the digestive tract. Many muscarinic receptors and fewer adrenergic (alpha 1 and beta 2) receptors are located in the smooth muscle of the digestive tract wall. Activation of the muscarinic receptors and blocking of adrenergic receptors leads to increased motility and relaxation of sphincters which augments material propulsion.

Postoperative ileus is a condition which sometimes results after surgery in which there is a disruption in the normal peristaltic activity of the GI tract. To re-establish normal gut motility, a muscarinic agonist such as **bethanechol** may be given to offer a "jump start."

For a summarized list of the effects of autonomic nervous activity, receptors in specific tissues, and drugs used to modify ANS activity, please review the tables below. You can also download a colored .pdf version of the table below.

A Table of Actions for the Sympathetic and Parasympathetic Divisions
Various Drugs Used to Modify the Actions of the ANS

10.3.1

A Table of Actions for the Sympathetic and Parasympathetic Divisions

Effector Organ	Sympathetic Effects (receptor)	Parasympathetic Effects (receptor)
Eye (iris)	Contraction of dilator pupillae muscles – mydriasis **(α1)**	Contraction of sphincter pupillae muscles – miosis **(M)**
Eye (ciliary muscle)	Relaxation **(β2)** for distant vision	Contraction **(M)** for accommodation of lens (near vision) and increase aqueous humor outflow into canal of Schlemm
Eye (ciliary body epithelium)	Increased aqueous humor production **(β1)**	—
Heart	Increased heart rate, increased force of contraction and increased conduction rate **(β1)**	Decreased heart and conduction rate **(M)**, decreased atrial contractility **(M)**
Arterioles(skin,abdominal viscera, kidney)	Strong vasoconstriction **(α1)**	—
Arterioles(skeletal muscle)	Weak vasoconstriction **(α1)** Vasodilation **(β2)**	—
Vessels(heart)	Vasoconstriction **(α1)**, Vasodilation **(β2)**	—
Lungs	Dilates Bronchioles **(β2)**	Constricts bronchioles **(M)**
Uterus, pregnant	Constriction **(α1)**, relaxation **(β2)**	Contraction **(M)**
Gastrointestinal tract wall	Decreased tone **(α1,α2,β2)**	Increased tone **(M)**

Gastrointestinal tract sphincter	Contraction **(α1)**	Relaxation **(M)**
Gatrointestinal tract secretion	—	Increased **(M)**
Kidney	Increased renin release **(β1)**	—
Bladder wall (detrusor muscle)	Relaxation **(β2)**	Contraction **(M)**
Internal urinary sphincter	Contraction **(α1)**	Relaxation **(M)**
Pancreas	Decreased insulin secretion **(α2)**, decreased exocrine secretion **(α)**	Increased insulin secretion **(M)**, increased exocrine secretion **(M)**
Fat cells	Lipolysis **(β3)**	—
Liver	Glycogenolysis **(α1,β2)**, Gluconeogenesis **(α1,β2)**	—
Piloerector muscles of skin	Contraction **(α1)**	—
Salivary gland	Constriction of vessels & small production of a viscous saliva. **(α1)**	Dilation of vessels & large production of thin saliva **(M)**
Sweat gland	Generalized sweating **(M)** Localized sweating(stress) – palms & soles **(α1)**	—
Adrenal medullae	Increased secretion of EPI or NE **(N)**	—

10.3.2

Various Drugs Used to Modify the Actions of the ANS

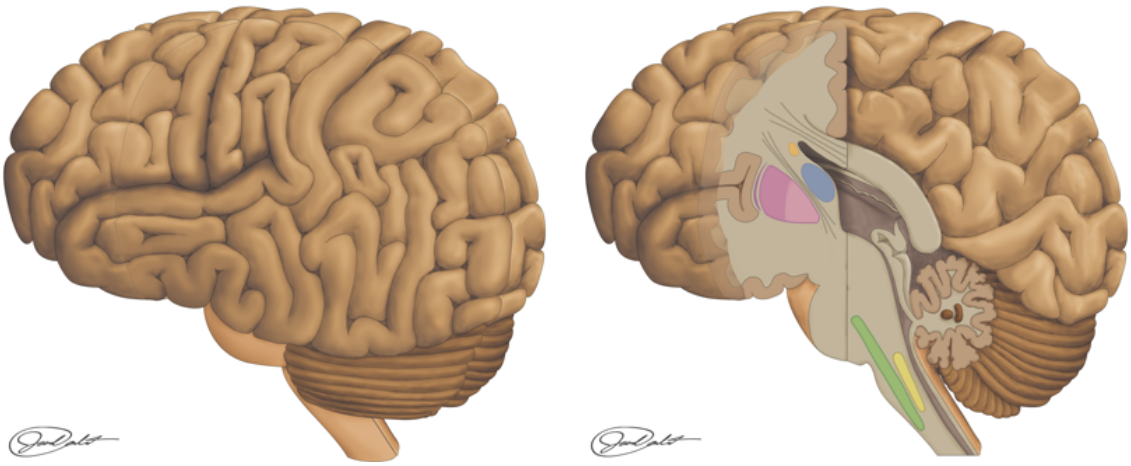
Drug Name	Receptor and antagonist/agonist	Indication(use)
Phenylephrine	α_1 agonist	Mydriatic
Atropine	Muscarinic antagonist	Mydriatic/emergency medicine to increase heart rate
Pilocarpine	Muscarinic agonist	Glaucoma
Betaxolol	Beta antagonist	Glaucoma
Prazosin	α_1 antagonist	High blood pressure, Raynaud's disease
Terazosin	α_1 antagonist	High blood pressure, hyperhidrosis
Atenolol	b_1 antagonist	High blood pressure
Ipratropium (Atrovent)	Muscarinic antagonist	Asthma/COPD
Albuterol	β_2 agonist	Asthma/COPD
Bethanechol	Muscarinic agonist	Postoperative ileus

11.0

MODULE 11: THE BRAIN

11.1

BRAIN OVERVIEW AND CEREBRUM

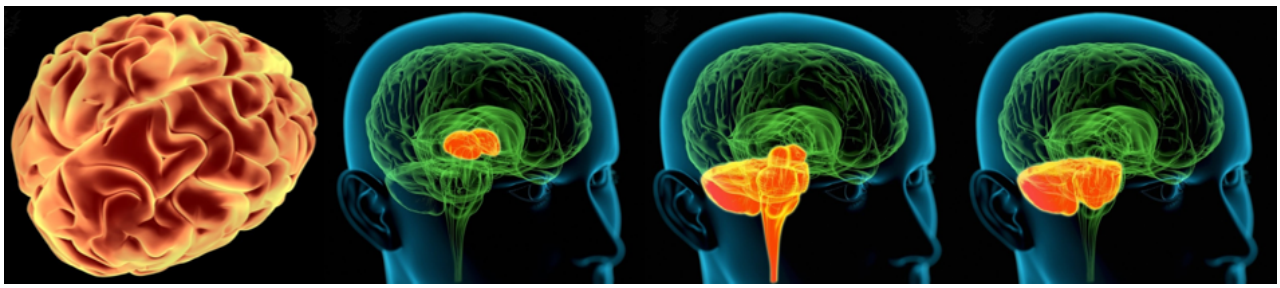


The Brain. *Image created by BYU-Idaho student Jared Cardinet Spring 2015*

The human brain is an incredible organ. In a Scientific American article published Oct 12, 2011, Mark Fischetti compared the human brain to the largest super computers. In the article he states: "For decades computer scientists have strived to build machines that can calculate faster than the human brain and store more information. The contraptions have won. The world's most powerful supercomputer, the K from Fujitsu, computes four times faster and holds 10 times as much data. And of course, many more bits are coursing through the Internet at any moment. Yet the Internet's servers worldwide would fill a small city, and the K sucks up enough electricity to power 10,000 homes. The incredibly efficient brain consumes less juice than a dim light bulb and fits nicely inside our head." The brain is the center of the nervous system and its functions include receiving, evaluating and responding to sensory input, storing information, planning future activities, reasoning and abstract thought. These functions are carried about by neurons, over 100 billion, which use action potentials as a means of communicating to other parts of the brain or to the rest of body. In order to generate the large number of neurons, our fetal brains had to produce about 500,000 neurons per minute during the early days of development. Each of us has about the same number of neurons in adulthood as they did at birth but the neurons grew and reached a maximum size at about age six. Synapses and brain organization continue through age 20 and perhaps beyond. How neurons send and receive action potential signals is well understood, but the way that neurons give rise to conscious awareness is still a mystery. It appears that a thought is a physical pathway of neurons in the brain. The more you use that pathway the easier it becomes to use, this is why repetition is so important in learning. This is also why having the thought "I'm so ugly," or "I'm a bad test taker" is never a good idea. Amazingly, aside from brain disease, your brain never loses the ability to learn and change. It is an urban legend that you only use 10% of the brain, we use it all!

As we study the brain we will identify different structures that are involved in specific functions. However, it is good to keep in mind that typically a given function will involve more than one region of the brain and that each region is probably involved in more than one function at a time. A good example is muscle movement. Normal, coordinated muscle movement involves several regions in the frontal lobe of the cerebrum, the basal nuclei and the cerebellum. We will see many other examples of this as we study the brain.

Although in appearance the human brain is nothing more than an oversized wrinkled walnut with the consistency of damp oatmeal, it can be divided into four major regions: the **cerebrum**, the **diencephalon**, the **brain stem**, and the **cerebellum** (see figure below). Additionally, each of these regions can be further subdivided into multiple structures. We will be learning the basic functions of each of these regions.



*© 2013 Encyclopædia Britannica, Inc. Downloaded from Image Quest Britannica; BYU-Idaho.***Diagrams, moving from left to right, illustrating the cerebrum, the diencephalon, the brain stem, and the cerebellum.**

The cerebrum is the largest and most superior part of the brain. It is highly developed in humans and separates man from all other species. It is divided into two **hemispheres** by the **longitudinal fissure** and each hemisphere is further divided into lobes. The classic division of the lobes is based on the cranial bones that overlay the cerebrum, hence there are four lobes, the **frontal**, the **parietal**, the **temporal**, and the **occipital** lobes. A fifth region, the **insula**, lies deeper in the cerebrum (Figure 2). Originally the lobes were designated solely based on their anatomical position but it is now known that each houses neurons with a specific function (more on this later). In order to increase the amount of surface area the cerebrum is arranged with numerous grooves and mounds. The grooves are called sulci (singular = sulcus) and the mounds are called gyri (singular = gyrus). Note: if the cerebrum were smooth it would have to be about the size of a beach ball to have the same amount of surface area. The arrangement of the gyri and sulci is fairly consistent between individuals but there are subtle individual differences. The outer 2-4 millimeters of the cerebrum is the **cerebral cortex**. It is composed of **gray matter**, which is made up of neuron cell bodies and dendrites. Under the gray matter is the medulla which is composed of **white matter**. White matter is primarily myelinated axons (Figure 3).The designation grey matter comes from the observation that it looks grey in fresh brain tissue and the inner layer looks white. Other clusters of cell bodies can be found deeper in the cerebrum within the white matter. These clusters are called **nuclei**. Recall that a cluster of neuron cell bodies in the peripheral nervous system is called a ganglion. The various nuclei of the cerebrum make up two important functional units, the **basal nuclei** and the **limbic system**.

Let's take a minute and try to explain how the brain works. Recall that typical neurons have three main components, a cell body, several dendrites and one axon. It is often helpful to compare the nervous system to a computer network. Using this analogy, each neuron cell body would be comparable to a computer or CPU. This is where information is stored, data is evaluated, and decisions are made. In a computer network individual CPUs are connected by fiber optic cables that send information from one CPU to another. The fiber optics would be analogous with the dendrites and axons in the brain. Their job is solely to transmit information, in the form of action potentials, from one cell body to the next. As the signal from one neuron reaches a synapse it generally results in one of two possible responses, it either excites the neuron, EPSP, or it inhibits the neuron, IPSP. It is the combination of these EPSPs and IPSPs that create the code that the nervous system uses. This is very much like the binary system of ones and zeros that computers use to process information. Hopefully, it is somewhat clear that it is the cell bodies (i.e. the gray matter in the cortex and

nuclei) that do all of the processing in the brain while the white matter simply carries messages from one neuron to the next. One last thing, in a computer network your CPU can connect to virtually every other CPU in the system. In the brain each neuron cell body is connected to up to 10,000 other neurons! It isn't hard to imagine how complex the brain circuitry must be.

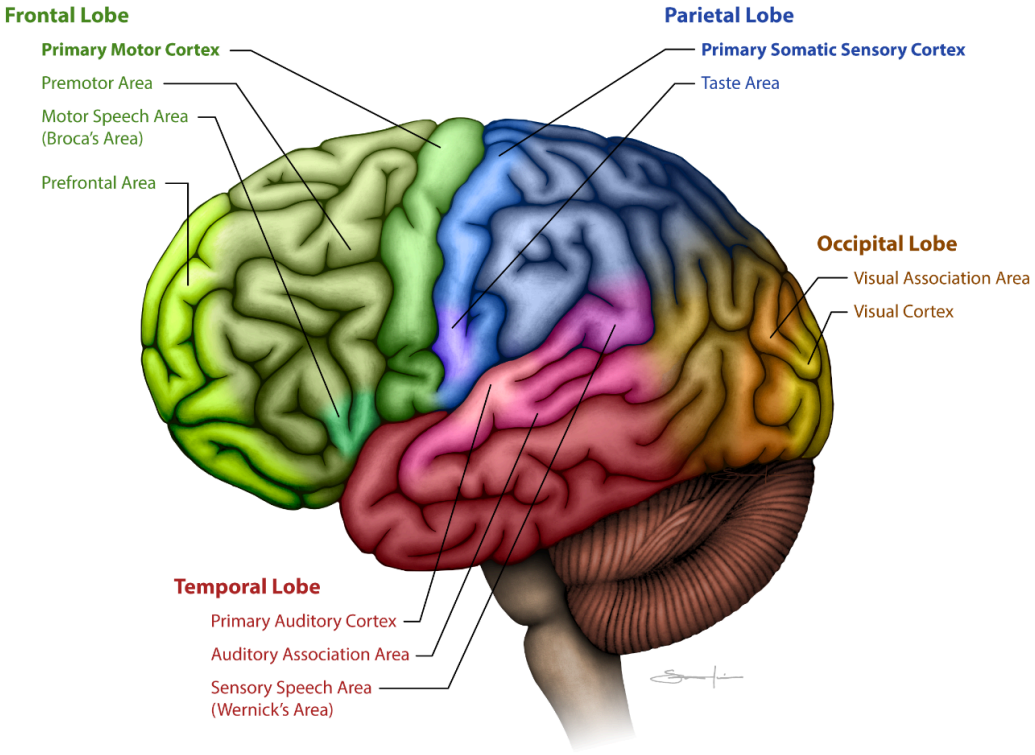
Recall that in our analogy of the computer network the axons function to transmit information like the fiber optic cables in a computer network. Functionally, the white matter within the medulla of the cerebrum can be divided into three types of fibers: **association fibers**, **commissural fibers** and **projection fibers**. Association fibers connect regions within a given hemisphere allowing the right frontal lobe to communicate with the right parietal lobe, etc. Commissural fibers allow the two hemispheres to communicate with each other, hence the right temporal lobe can talk to the left temporal lobe. Commissural fibers cross from one hemisphere to the other through an area called the **corpus callosum**. Projection fibers connect the cerebrum with other parts of the brain and to the spinal cord allowing information to be sent both out of and into the cerebrum.

Cerebral Cortex

11.1.1

Cerebral Cortex

The functions of the cerebral cortex include memory, attention, perception, thought, movement, language and consciousness. In other words, it allows us to be aware of ourselves, to remember names, to communicate to others and to move voluntarily. It contains billions of neuron cell bodies and dendrites, glial cells and blood vessels. Some specific functions of the cerebral cortex can be associated with the different lobes. For example, the frontal lobe is associated with motor behavior, the parietal lobe with processing and perception of sensory information. The occipital lobe is visual processing and perception and the temporal lobe processes hearing, vision, balance and language. Because of the vast array of functions, the cerebral cortex can be further divided into three generalized areas: **motor** areas, **sensory** area, and **association** areas.



Locations of Sensory and Association in the Brain. *Created by BYU-Idaho student, 2013.*

Motor Areas

One key function of the cerebrum is control of skeletal muscle. The motor areas of the cerebral cortex control voluntary movement and are localized in the frontal lobe, they include the **prefrontal cortex**, the **premotor cortex** and the **primary**

motor cortex. Decisions to perform a specific motor function originate in the prefrontal cortex (we will discuss other functions of this region later). Once the decision is made, information is sent to the premotor cortex. This is the staging or programming area. It must be determined which muscles will contract, what is the order of contraction, how much force each must generate, etc. for the desired movement. Once programmed, signals are sent to the primary motor cortex, which then relays the signal to the spinal cord via the upper motor neurons. It should be noted that this is a simplification of the process. For example, not all upper motor neurons originate from the primary motor cortex but may also come from the premotor area, or even the somatosensory cortex of the parietal lobe.

Anatomically the prefrontal cortex is the anterior-most region of the frontal lobe. Moving posteriorly the premotor cortex comes next. Finally, the primary motor cortex is the most posterior region of the frontal lobe. The frontal and parietal lobes are separated by the central sulcus. The gyrus just anterior to the central sulcus is the **pre-central gyrus**. This gyrus houses the primary motor cortex. Neurobiologist have mapped the primary motor cortex based on which parts of the body they control and found that primary motor cortex is organized in a toe to mouth arrangement. That is, the neurons that control the lower parts of the body are at the top of the pre-central gyrus while those that control the upper parts of the body are located at the bottom of the pre-central gyrus. Additionally, the neural areas are not proportional in size in the body part that they control. The next image shows the body parts superimposed on the pre-central gyrus drawn in proportion amount of brain tissue devoted to controlling them. It becomes obvious that those areas responsible for the most motor units control fine, intricate movement. The hands for example, have much more of the primary motor cortex devoted to them than areas that principally produce more gross movements like the legs and torso. It should be noted that this map of the body is not cleanly segregated and contains a considerable amount of overlap. Studies have shown that even a single neuron in the primary motor cortex can influence the activity of multiple muscles related to multiple joints. Still, the map does provide a starting point.

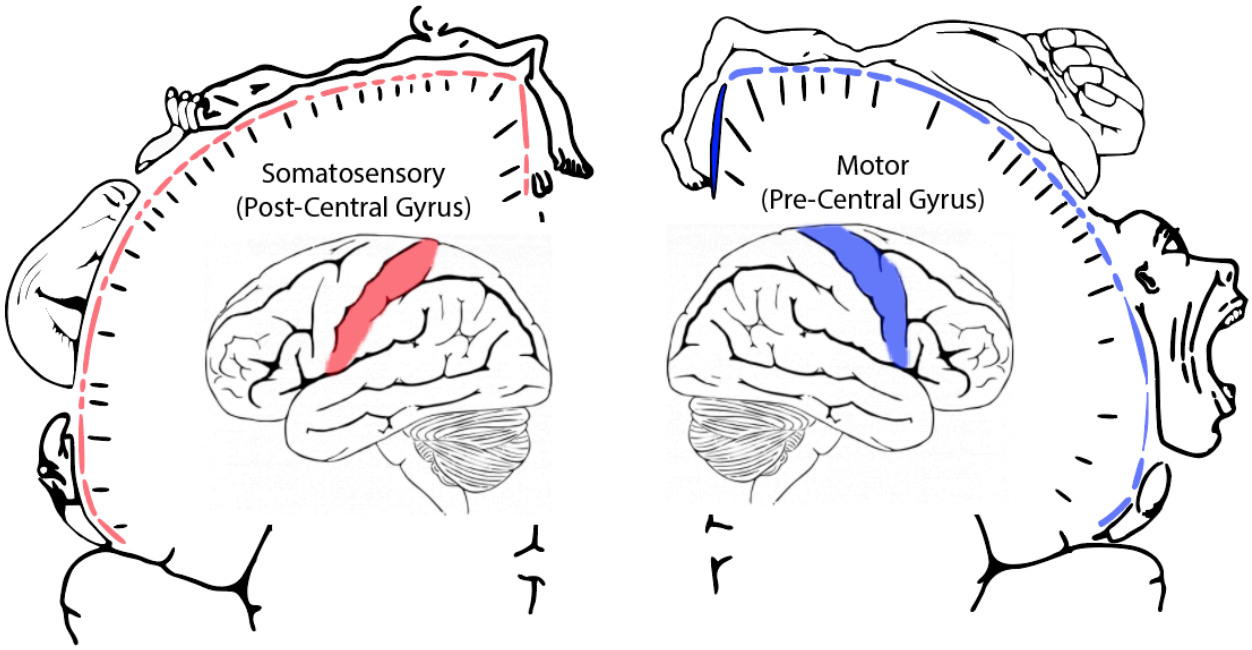


Image created using public domain images from the following sources
https://commons.wikimedia.org/wiki/File:Homunculus-ja.svg; http://cliparting.com/free-brain-clipart-5051/.

Image above represents the proportion of the primary motor cortex and the somatosensory cortex that controls different regions of the body. The hands have the largest region of the brain devoted to motor control. The face and hands receive a large amount of sensory information. A map like this is called a homunculus (or topographic diagram).

Sensory and Association Areas

For each of the different senses, there is a region of the cortex designated for perceiving that sense. In figure 5 several of these primary sensory cortexes are labeled.

The primary somatic sensory cortex is located in the postcentral gyrus of the parietal lobe. This region receives sensory information from the skin and from the proprioceptors. It is therefore responsible for perceiving our senses of touch, pressure, temperature and pain, as well as informing us about the position and movement of the body. Note that the word somatic or soma means body. As is seen with the primary motor cortex, the sensory positioning of the body is also represented upside down on the primary somatic sensory cortex. Likewise, the size of the area in the brain devoted to a particular area of the body is dictated by the density of sensory receptors. For example, fingers and lips have large areas in the somatosensory cortex devoted to them compared to the torso or legs. If you want to feel the texture of an object you typically use your fingertips rather than rubbing the object on your leg! The primary visual cortex is located in the occipital lobe. It receives input from the eyes and generates images from the input it receives. The primary auditory cortex is located in the temporal lobe and converts signals coming from the ears into sounds. Not shown on the image is the primary olfactory cortex (smell) which is located at the junction of the temporal and frontal lobes (recall that the division of the cerebrum into lobes was based solely on anatomical position and not on function). Also not shown is the primary gustatory cortex (taste) which is found in the boundary between the insula and frontal lobes. Two other sensory cortexes, the primary visceral cortex and the primary equilibrium cortex are found in the insula.

Near each of the primary sensory cortexes is a sensory association area for that particular sense. As the name implies, these regions help us associate what we are currently sensing with our past experiences. For example, when you see a face after the image is perceived in the primary visual cortex it is sent to the visual association area for recognition. Do I know this person, have I seen this face before, is this person safe or are they a threat? The association areas allow us to make sense of what we are experiencing and react appropriately.

Executive Functions

We have been discussing the motor and sensory functions of the cortex, but all animals possess these regions. An important difference between humans and all other species is the development of the prefrontal cortex. This region is well developed only in primates and particularly in humans. It is this part of the brain that is responsible for what have been called executive functions. These include planning, reasoning, abstract thought, self-control, decision making, differentiation between good and bad, between better and best, and in understanding consequences of our actions. In addition, it is thought that our personalities are determined by this region as well as the storage of short-term or working memory. Interestingly, this is the last region of the brain to fully develop and mature. It has been suggested that full development isn't completed until our late teens or early twenties. Did you ever wonder why you did that really stupid thing when you were a teenager? It is likely that at the time you were not completely capable of connecting the action with the consequences. Not that you can use this as an excuse but maybe we should be a bit more patient with children and teenagers, realizing that it may be beyond their capability to think the same way you do.

Lateralization of the Hemispheres

As mentioned above the commissural fibers of the white matter allow the two hemispheres to communicate with each other. Many functions, like movement and sensations, are carried out equally between the two hemispheres. It is true that our right hemisphere controls the left side of our bodies and our left hemisphere controls the right side of the body. Sensory information is usually shared between the two hemispheres, which allows for some interesting things to happen, like depth perception and localizing the origin of sounds. A few functions, however, seem to be restricted to one hemisphere or the other. Speech, for example, is a left hemisphere function in most people (there are always exceptions). There are two important cortical regions found in the left hemisphere that do not have counterparts in the right hemisphere. These are **Broca's area** and **Wernicke's area**. Broca's area is located in the frontal lobe (figure above) and is the motor speech area. The ability to speak and write is associated with this region. Damage to Broca's area results in the inability to speak or to write clearly. This is called “expressive aphasia”. Wernicke's area is located in the posterior temporal lobe (actually exists in an area where the temporal and parietal lobes meet) and is required for

understanding both spoken and written words. Damage to this region results in the inability to understand spoken or written words. This is called “receptive aphasia”. If you were asked to read this paragraph out loud the following sequence of events would have to happen. Information from the eyes would reach the primary visual cortex where they would be perceived as words. This information would be sent to the visual association area for recognition of the words. A signal would then be sent to Wernicke's area for the words to be understood. Wernicke's area would then send a signal to Broca's area where the words are formulated to be spoken. Broca's area then sends the information to the premotor cortex for programming. Once the motor activity is programmed, it is sent to the primary motor cortex which relays the signals to the muscles involved in generating speech.

Other functions that seem to be more prominent in one hemisphere than the other are analytical skills, which like speech seems to be a left hemisphere function in most people. Similarly, spatial perception and musical ability are more right hemisphere functions in most people. One interesting theory was that left-handed individuals, who are right brain dominant in their motor function, should be more artistically inclined since the right hemisphere is considered the artistic side of the brain. However, studies indicate that handedness does not increase or decrease the likelihood of someone being artistically talented.

11.2

THE DIENCEPHALON,BRAINSTEM AND CEREBELLUM

The diencephalon (interbrain) is a region at the core of the brain and is surrounded by the cerebral hemispheres. It is the connection between the brain stem and the cerebrum and consists of three gray matter structures; the **thalamus**, the **hypothalamus** and the **epithalamus** (see figure below).

11.2.1

The Thalamus

By the simplest definition, the thalamus is the sorter or relay center for information coming into the cerebral cortex from all parts of the body (sensory impulses). With the exception of smell, afferent neurons from all parts of the body converge and synapse in the thalamus which in turn relays the information to specific regions of the cerebral cortex. The thalamus edits and sorts out information and then categorizes similar functions to be relayed as a group to the appropriate areas of the cerebral cortex. Thus, specific localizations and interpretation of stimuli occur in the cerebral cortex but only after careful sorting through the gate keeper, the thalamus. A recent study suggests that the thalamus plays an important role in our ability to concentrate on the task at hand by ignoring distracting sensory input. (Wimmer et al. "Thalamic control of sensory selection in divided attention," Nature, October 21, 2015. DOI: 10.1038/nature15398). It also plays an important role in regulating out states of sleep and wakefulness. In addition, it is thought to play an important role in maintaining the aroused state and damage to the thalamus can result in coma.

11.2.2

The Hypothalamus

The hypothalamus is so named because of its position below the thalamus. The hypothalamus is the visceral control center, it regulates functions of the internal organs. As such, it is chiefly concerned with maintaining homeostasis. Due to the key role it plays in maintaining normal body function it is sometimes referred to as the brain within the brain. Some of the important functions of the hypothalamus are listed in the table below.

Table: Select homeostatic roles of the hypothalamus.

The autonomic nervous system	The autonomic nervous system is a system of neurons that automatically regulate function such as heart rate, blood pressure, digestion, etc. (functions we don't have to consciously think about). The hypothalamus regulates many of the activities of the activity of the autonomic nervous system by controlling centers in the brain stem and spinal cord.
Emotions	Hypothalamic neurons are involved in the perception of pleasure, fear and rage.
Body temperature	Select groups of hypothalamic neurons monitor blood temperature directly as well as respond to inputs from other thermoreceptors throughout the body and then send appropriate signals to systems that help regulate body temperature such as sweat glands.
Food intake	Select neurons respond to blood levels of nutrients and regulate feelings of hunger or satiety.
Water balance	Osmoreceptors (modified neurons) in the hypothalamus respond to changing salt concentrations in the blood which in turn elicit responses from the kidneys as well as regulate thirst.
Sleep	Neurons in the hypothalamus have been linked to our biological clock.
Endocrine system	The hypothalamus regulates the pituitary gland. A major endocrine organ gland that regulates numerous body functions including metabolism and reproduction.

11.2.3

The Epithalamus

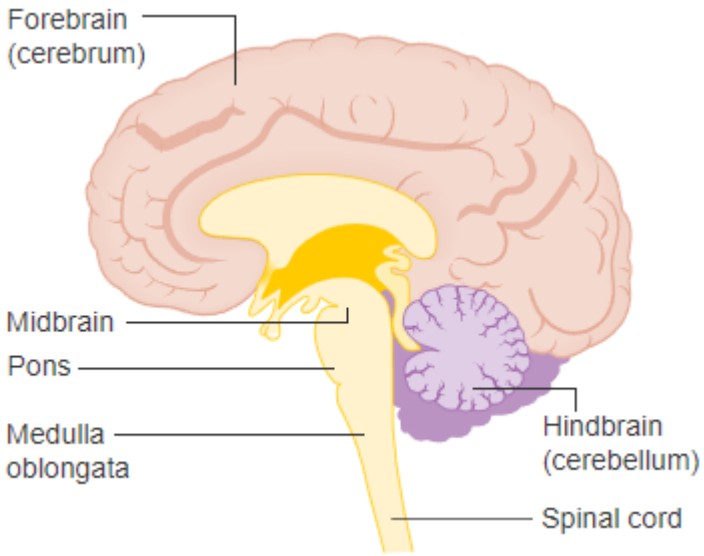
The epithalamus is the most dorsal of the structures of the diencephalon. Within the epithalamus are several important structures including the habenular nuclei and the pineal gland (sometimes called the “pineal body”). The habenular nuclei have been shown to have involvement in several limbic system type functions including negative reward processing.

The pineal gland secretes the hormone melatonin in response to the light dark cycle (melatonin secretion is stimulated by the dark and inhibited by light). Melatonin has been implicated in the regulation of our sleep patterns and in regulating reproduction in seasonal breeding animals (see endocrine modules). Melatonin may play an important role in puberty as pineal tumors have been linked to the onset of precocious puberty. The habenular nuclei are thought to be involved in pain processing, reproductive behavior, learning, sleep-wake cycles, stress responses and nutrition. Many of these functions are known to be related to the limbic system, indeed, the epithalamus is considered the bridge between the limbic system and the cerebrum (more on the limbic system later).

11.2.4

Brainstem

The region of the brain that connects the brain to the spinal cord is the brain stem. The brain stem is subdivided into three regions: the **midbrain**, the **pons**, and the **medulla oblongata**.



Brain Stem. *By Cancer Research UK [CC BY-SA 4.0 (http://creativecommons.org/licenses/by-sa/4.0)], via Wikimedia Commons*

The brain stem is also the site where groups of axons (nerve tracts) either exit the brain as cranial nerves or continue on into the spinal cord. Indeed, ten of the twelve pairs of cranial nerves (III - XII) exit the central nervous system from the brain stem. It should also be noted the brain stem is essential for maintaining critical body functions, such as respiration and regulation of the heart, and that we cannot survive without its functions. We can survive without a cerebrum, although we would not be conscious that we were alive, but the body cannot survive without the brain stem.

The Midbrain

The midbrain is the upper most portion of the brain stem and is situated between the diencephalon and the pons. Running through the midbrain is a hollow tube which connects the third and fourth ventricles (The ventricles of the brain are hollow spaces filled cerebral spinal fluid). Three unique clusters of cell bodies (nuclei) are observed in the midbrain; the **corpora quadrigemina**, the **substantia nigra**, and the **red nucleus**.

Spend time looking at the image search on the midbrain.

The corpora quadrigemina are subdivided into two regions, the two superior colliculi and the two inferior colliculi. The paired **superior colliculi** (colliculus; singular) coordinate the movement of our eyes as we track a moving object. This reflex involves input from the eyes via the optic nerves, integration of the signal in the superior colliculi, and efferent signals to the muscles that control eye movements via cranial nerves III, IV and VI. The **inferior colliculi** help coordinate head and eye movements in responding to sudden sounds that cause you to abruptly move your head and turn your eyes toward the sound. For example, “Jaws,” or “Watcher in The Woods” kind of stuff.

The substantia nigra (dark substance) gets its name from its dark appearance in fresh tissue. This is due to the pigment neuromelanin (similar to melanin) which is produced in these cells. The neurons from this region produce dopamine as their neurotransmitter and neuromelanin which is derived from the same precursor that produces melanin, L-dopa. Neurons from the substantia nigra ascend to the cerebrum and synapse with structures of the basal nuclei, a part of the brain involved in skeletal muscle control. Degeneration of these neurons is the cause of Parkinson’s disease, a condition in which the patient is unable to suppress unwanted muscle contractions resulting in constant tremors of the extremities. Parkinson’s disease also results in some muscles being overly “stiff” and others having too little tone. Parkinson’s disease makes it difficult to coordinate voluntary and involuntary contractions.

Neurons in the red nucleus contain inordinate amounts of iron which when oxidized conveys a red hue. In many vertebrates, it is a relay center for motor pathways that affect limb flexion. It is thought that in humans; the red nucleus influences arm swing during gait, crawling in babies and motor control of some of the larger shoulder and arm muscles, but not the legs.

The Pons

The pons region of the brain stem contains nuclei that contribute to the control of sleep, respiration, swallowing, bladder control, hearing, equilibrium, taste, eye movement, facial expressions, and posture. It may also play a role in generating dreams. In addition, the pons (which means bridge) connects the cerebellum to the cerebrum.

The Medulla Oblongata

"Alligators are ornery cause of their Medulla Oblongata"! (Water boy, 1998). Actually, the medulla oblongata probably has nothing to do with ornery … surprise! As if movies ever tell the truth … Anyway, the medulla oblongata has a crucial role in body homeostasis. It has beens said that the medulla oblongata controls many of the vital reflexes for life. Neurons in the medulla oblongata adjust the force and rate of heart contractions, they generate and modify the depth of breathing and regulate other fun activities like vomiting, hiccupping, coughing and sneezing. Additionally, the medulla is where most of the neurons that control voluntary skeletal muscle contraction cross over to the opposite side of the brain stem. The result of this crossing over is that the right side of the brain controls muscles on the left side of the body and the left side of the brain controls muscles on the right side of the body. The structure in the medulla where this crossing over is the olives of the medulla. The technical name for crossing over is **decussation**. This word comes from "deca" the prefix for the number 10 and the Roman Numeral for 10 which is X. The symbol X implies crossing over.

11.2.5

Cerebellum

The cerebellum ("little brain") sits under the occipital lobes of the cerebral hemispheres and attaches to the brain stem. Its structure is similar to the cerebrum in that it has a cortex composed of gray matter with white matter in the center. Likewise, the surface of the cerebellum is composed of folds called folia. It functions in motor learning, motor coordination and equilibrium. Smooth, coordinated skeletal muscle movement requires a functioning cerebellum. As we practice and learn complex movements the cerebellum is crucial in fine tuning the movements. When the motor cortex of the frontal lobes sends orders to the muscles to perform a particular task a copy of those orders is sent to the cerebellum. It, in turn, receives feedback from the proprioceptors in the muscles and joints as well as information from the inner ear relating to balance and equilibrium and compares what is actually happening with what the motor cortex ordered. Based on these comparisons, information is relayed back to the motor cortex in the cerebrum to fine tune the motor activity. The end result is the development of smooth, coordinated movements and agility for tasks like typing, driving, piano playing, dancing etc. In addition, procedural memories for tasks like how to walk or ride a bike are stored here. Other studies using brain imaging and observations of patients with cerebellar injuries suggest that the cerebellum also plays roles in language, thought processing, and emotions.

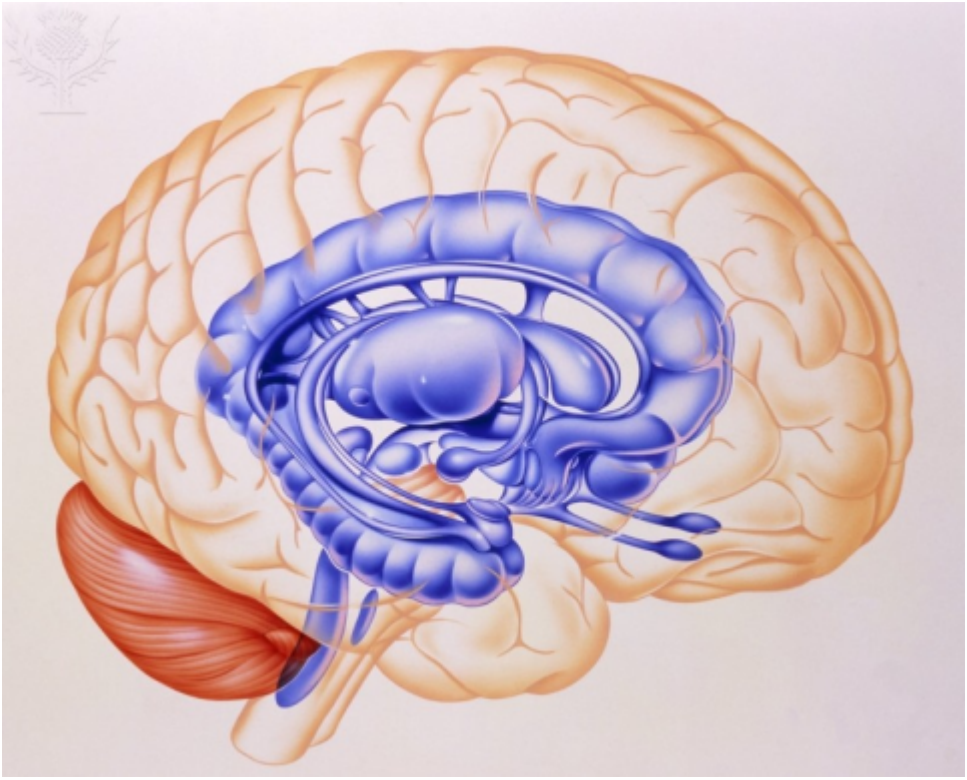
11.3

THE LIMBIC SYSTEM, BASAL NUCLEI AND RETICULAR ACTIVATING SYSTEM

11.3.1

The Limbic System

The limbic system is probably the most primitive region of the cerebrum and hence is involved in some of the basic survival functions of the brain, namely; memory, reproduction and nutrition. This system is composed of several nuclei (clusters of cell bodies), the most conspicuous of which are the hippocampus, the cingulate gyrus and the amygdala. The image below highlights the components of the limbic system in blue. It is linked to the diencephalon and most of the output from the limbic system passes through the hypothalamus. As such it is closely tied to the autonomic nervous system. The limbic system functions as our emotional center, and yes, guys have a limbic system as well. As mentioned above, the limbic system also plays a role in the formation of memories. It is thought to play a role in remembering where food is found and the pleasure associated with eating. Additionally, it remembers the pleasure from sex as damage to this area results in voracious appetites for food and/or sex. The functions of two important structures of the limbic system, the hippocampus and the amygdala are explained below.



Limbic System (in blue). *© 2013 Encyclopædia Britannica, Inc. Downloaded from Image Quest Britannica; BYU-Idaho.*

Hippocampus

The hippocampus is involved in various processes of cognition and spatial memory, or the what, when, and where of memory. In addition, the hippocampus is very important in learning. In fact, neurologists have observed increases in

neuron numbers in response to training of a task which ultimately resulted in vast improvement in the learning of the task.

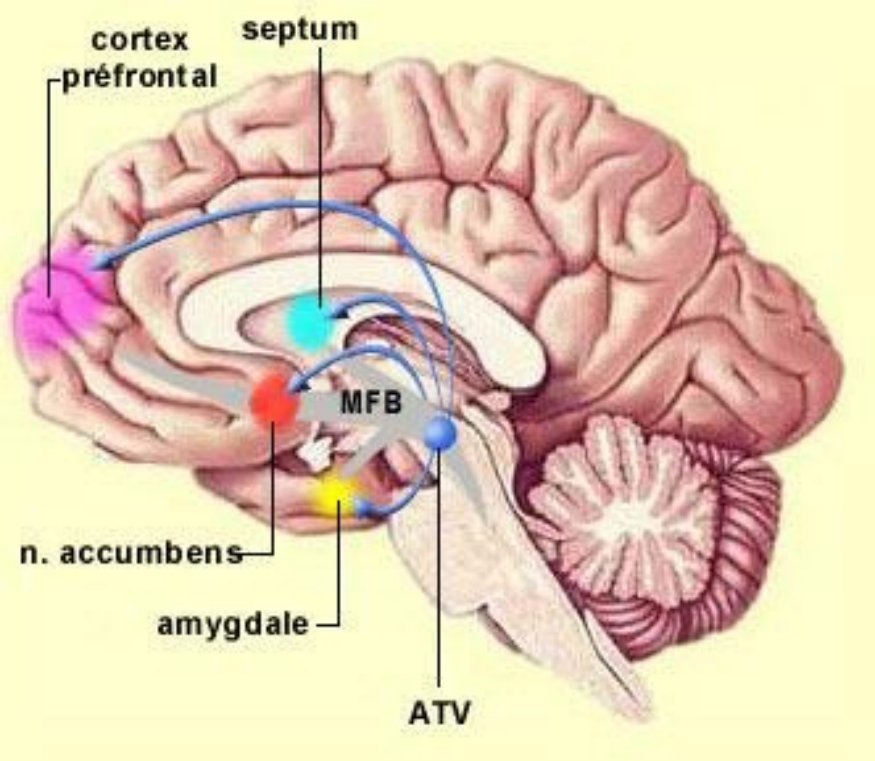
Amygdala

The amygdala is also involved in many cognitive processes and memory but in contrast to the hippocampus the amygdala functions in episodic and autobiographical types of memories. In addition, the amygdala is important in maintaining attention. In fact, if you have gotten this far in your reading you are probably not amygdala challenged. A function that appears to be specific to the amygdala is social processing, in particular facial recognition and the evaluation of first impressions. The amygdala is also involved in "fear conditioning". Stimuli that we might regard as frightening trigger the amygdala to help us form a fear memory. If similar stimuli occur in the future, the amygdala will help us recall the memory and experience fear again.

The Reward Center

Because addictions are becoming such a problem in modern society we have added the following section to describe the physiology of addiction.

The reward pathway consists of five brain structures: the nucleus accumbens, the ventral tegmental area (also called area tegmental ventral or ATV in the picture below), the amygdala, the septal nuclei (septum), and the medial forebrain bundle (MFB).



Reward Centers of the Brain. *Downloaded from Wikimedia Commons Dec 2013; Author: lecerveau.mcgill.ca; Licensed under the Creative Commons Attribution-Share Alike 3.0 Unported*

The Nucleus Accumbens

The nucleus accumbens is a collection of neurons that play major roles in reward, pleasure, laughter, addiction, aggression, fear, and the placebo effect. In short, the nucleus accumbens is very important for motivation and pleasure. This structure has also been linked to addiction and depression as damage to the nucleus accumbens results in lack of motivation and loss of addictive behaviors. The nucleus accumbens uses dopamine and serotonin as the preferred

neurotransmitters. In response to dopamine secretion the nucleus accumbens elicits feelings of pleasure while serotonin release has a calming influence. The surge of these neurotransmitters during addictive behaviors triggers the neural activity that is correlated with the sensation of reward. Studies have shown that when individuals crave a substance, their neural activity increases in anticipation of the future pleasure.

The Ventral Tegmental Area

The ventral tegmental area is a collection of cell bodies that release the neurotransmitter dopamine (dopaminergic) in response to rewards. It is important in cognition, motivation and intense love emotions. The neurons project to numerous areas of the brain, in particular the nucleus accumbens. Neurons in the ventral tegmental area respond to novelty, unexpected rewards and rewards that are predictive (rewards that the brain has previously learned are rewarding). Under resting conditions the dopaminergic neurons are phasic, with release of dopamine occurring at predictable consistent rates. However, when a stimulus is received the neurons send multiple action potentials to the nucleus accumbens which results in the increased release of dopamine. The artificial release of dopamine by the ventral tegmental neurons can occur in response to heroin, cocaine, alcohol, opiates, marijuana, nicotine and amphetamines. The effect of these drugs is to prolong the action of dopamine on the nucleus accumbens. Repeated use of drugs results in functional changes within the reward circuitry induced by both molecular and cellular adaptations. In response to the functional changes, the body adapts to the increase in dopamine release so that normal release of dopamine is not sufficient to reward the brain. In response, the body enters a state where the drugs become necessary to restore the normal homeostatic state. In animal studies, even after the final stages of withdrawal have passed, an organism will "re-enter" the addictive state if the exposed again to the drug.

Amygdala

As already stated the amygdala neurons are involved in the formation and storage of memories associated with emotional events. Studies have shown that the greater the emotional arousal that occurs with a learning event the greater the retention. The amygdala is also associated with social interactions. Loss of function of the amygdala is associated with loss of fear and the loss of the ability to discriminate between animate and inanimate objects. It is interesting to note that homosexual men have more female-like patterns in the amygdala than heterosexual men and homosexual females show more male-like patterns in the amygdala when compared to heterosexual women. The amygdala is involved in the processing of personal space. The amygdala is also associated with sexual and aggressive behavior.

The Septal Nuclei

The septal nuclei are very similar in function to the nucleus accumbens, both playing a role in the reward and reinforcement pathways. The septal nuclei differ from other areas in that the signals sent to the amygdala from the septal nuclei are inhibitory. For example, activation of the amygdala can result in sexual behavior and the desire for physical contact. This activation of the amygdala can then be modified by the septal nuclei through inhibitory signals to allow the person to decipher what is appropriate and what is not. Some researchers suggest that this modulation is essential in forming closer more long-lasting emotional bonds between partners.

The Medial Forebrain Bundle

The medial forebrain bundle serves to carry information between the ventral tegmentum and the nucleus accumbens. It is no wonder that humans report that stimulation of this area of the brain is intensely pleasurable.

Together, these five structures are called the reward system and are very important for driving our feelings of motivation, reward and behavior. These feelings include those that are necessary for the survival of the person such as food, sexual contact, and protection. The reward pathway is a natural and important component of our enjoyment in life. However, addictions occur when the reward system is abused.

An **addiction**, by definition, is the continued use of a mood or behavior altering substance despite adverse

consequences. They result from the motivated repetition of the same thougths and behaviors unitil they become habitual. Addictions can include, but are not limited to, drug abuse, sexual addiction, gambling, overeating and even exercise addiction. **Dependence** upon addictions occurs when the body learns to adjust to the substance by incorporating it into the body's normal function. This adjustment creates conditions of tolerance and withdrawal. **Tolerance** refers to the body's ability to adapt to the substance which then requires increasing amounts of the substance to achieve the original effect. **Withdrawal** refers to the symptoms, both physical and psychological, experienced when the substance use is discontinued or even reduced. Withdrawal symptoms include anxiety, irritability, intense cravings, nausea, headaches tremors and hallucinations.

The "addictiveness" of a substance is often determined by one of four factors:

1. The substance is a highly stimulating version of a natural product (for example; high-calorie foods)
2. The substance is available in limitless supply
3. The substance comes in lots of varieties (novelty)
4. The substance causes us to binge without realizing it is triggering brain changes

Very addictive substances usually meet 2 or more of the categories (high-calorie foods, exercise) or even all 4 categories (internet porn). All addictions have a common theme; they induce physiological changes in certain structures of the brain. So why do addictive patterns occur? Addictions arise because of the *misuse* of the normal reward pathways of the brain.

Internet Pornography

Now that we have a basic understanding of the reward pathway, we can attempt to explain why some substances can become so incredibly addictive. As previously stated, it is important to understand that the reward circuitry is important even necessary for survival. At the forefront of the reward system is the neurotransmitter dopamine. In brief, the purpose of dopamine is to motivate, thus, the bigger the dose, the bigger the motivation and the more desire to do something. For example, chocolate and ice cream are great motivators for dopamine release, broccoli, by comparison, is not. Not surprisingly, sexual activity causes large surges of dopamine. Dopamine surges increase with novelty. Internet pornography is especially enticing because endless novelty is just clicks away (points 2 and 3 of "addictiveness"). The brain will eventually adapt, and require access to a super-stimulating reward to maintain normal homeostasis. This adaptation is what leads to addiction.

All addictions lead to the same major brain changes:

1. **Desensitization:** the neurotransmitter dopamine declines and the dopamine receptors are down regulated. This creates a less sensitive area for natural dopamine release and leaves the individual craving for activities that result in high dopamine release. The addict will tend to neglect interests and/or behaviors that were once of high personal value.
2. **Sensitization:** The newly wired brain and reward pathway will start to turn on in response to any addiction-related stimuli or even thoughts.
3. **Hypofrontality:** The frontal lobe centers, those associated with understanding consequences, begin to weaken. The result is a reduced response to the ability to foresee consequences of actions.
4. **Dysfunctional stress circuits:** This means that stress, which before the addiction was easily managed, can now trigger relapses when the individual comes under stress.

In addition to dopamine, another protein called DeltaFosB serves to modulate many of the activities of the reward circuitry. Continued over-consumption of natural rewards (sex, sugar, high-fat, exercise) causes DeltaFosB to accumulate in the reward circuitry. DeltaFosB is also a protein that motivates, but unlike dopamine, the simple purpose of DeltaFosB is to "take while the taking is good". In other words, it is a primitive binge mechanism already in place for natural rewards. Think of our ancestors who didn't have complete access to fruit. This circuitry allowed them to identify and consume the fruit before it was lost. This would be a very beneficial reward when there was very little to be had, however, when there is plenty it makes addictions to things like junk food and internet pornography extremely easy.

Perhaps Dr. Gary Wilson put it best when he said; "nerve cells that fire together, wire together." This re-wiring and overproduction of the neurotransmitter dopamine and the protein DeltaFosB strengthen connections between the nerve cells and make it easier for those particular neurons to communicate. Habitual viewing of internet pornography results in addiction-related brain changes. Thus, just as water flows down the path of least resistance, so do the impulses, and ultimately our thoughts and desires. Thus, normal everyday pleasures are viewed as boring because they do not stimulate those pathways, but anything associated with the pornography addiction results in hypersensitivity.

Heavy pornography use can also result in tolerance unless the user moves into new directions in search of more intense experiences to produce a more powerful chemical response. Therefore, the more intense the event, for example adding masturbation, the stronger the response, which results in an increased brain re-wiring.

What makes Internet pornography unique and so addictive?

1. **It affords extreme novelty:** Internet pornography allows for hundreds of "new" scenes per session. Novelty is highly stimulating. Today's pornography is even more addictive than pornographic magazines of the past. With Internet pornography, one can escalate both with new scenes and with new types of pornography. It's quite common for a user to move to ever more extreme and degrading forms of pornography.
2. **Limitless exposure:** Unlike food and drugs, in which there is a physical limit to consumption, there are no physical limitations to Internet pornography consumption. The user can simply click and begin the process all over again.
3. **Lack of Aversion mechanism:** An aversion system is activated when you don't like the symptoms, for example, eating too much is associated with pain so typically you stop bingeing in response to the pain (aversion system). Internet pornography doesn't have any immediate side effects to activate the natural aversion system. However, long term side effects include:

    a. Distress about escalation to more extreme pornography

    b. Frequent masturbation

    c. Uncharacteristic, worsening social anxiety or lack of confidence

    d. Morphing pornography tastes that don't match sexual orientation

    e. Inability to concentrate, and extreme restlessness

    f. Depression and anxiety

Many of these symptoms arise from the changes in dopamine levels and dopamine receptors.

Is there hope for the internet pornography addict?

Fortunately the brain, including the reward center, is very plastic in its ability to restore order. The first step in addiction recovery is the re-balance of the brain. This begins with stopping all sexual stimulation, including pornography, masturbation, fantasies, chat rooms, erotic stories or surfing on the internet. Other triggers must be carefully watched, these include boredom, loneliness, anger, stress and being overly tired. The pathways must be rebuilt to respond to natural stimulants. Additionaly, intense shame and guilt are more likely to thwart recovery than facilitate it.

The Church of Jesus Christ of Latter Day Saints provides many resources for learning how to prevent or recover from a pornography addiction and other addictions by focusing on the power of Christ’s Atonement. Here is the link to their websites:

- https://books.byui.edu/-hAkt
- https://books.byui.edu/-wAWS

11.3.2

The Basal Nuclei

Another key group of nuclei found in the cerebrum are basal nuclei (also referred to as the basal ganglia). Key nuclei of this group include the caudate nucleus, the putamen and the globus pallidus. The basal nuclei also receive input from the substantia nigra of the midbrain. The main function of the basal nuclei is in regulating motor control. Although the precise details are not fully understood, it seems to play a key role in preventing incorrect and/or inappropriate movements. It seems to be key in regulating what are referred to as stereotyped movements such as swinging the arms when walking. Additionally, it is thought to play a key role in initiating, stopping and monitoring the intensity of voluntary motor movements. Output from this system does not go to the muscles themselves, rather output is sent to the motor centers in the frontal cortex where adjustments and corrections can be made to the outgoing signals. Damage to the basal nuclei can result in conditions that produce excessive movement (Huntington's disease) or too little movement (Parkinson's disease).

11.3.3

The Reticular Activating System

The reticular activating system (RAS) is a diffuse network of neurons in the brain that interact with structures such as the hypothalamus, thalamus, cerebral cortex and the cerebellum. It includes the **reticular formation** that originates in the brain stem near the pons and radiates into the cerebrum. Functionally this network of neurons is called the name reticular activating system. The primary function of this system is to maintain the brain in a state of alertness or arousal. It is also involved in regulating our sleep wake cycles. Damage to this system results in the inability to remain awake and alert. Narcolepsy, for example, is thought to be due to malfunctioning of the RAS. If the system is suddenly shut down, like from a blow to the head, the victim will lose consciousness. Caffeine has the effect of exciting the reticular formation fibers so that one "feels" more awake or as if they have more energy. Likewise bright light, cold water on the face and noxious chemicals (smelling salts) stimulate the RAS while dim lights, soothing music, warmth and general anesthetics suppress this system.

11.4

HIGHER BRAIN FUNCTIONS: THE EEG, SLEEP AND LEARNING

A characteristic of all excitable tissues is that they are capable of generating and propagating signals that involve changes in the electrical charge on the cell membrane. We have described these changes in earlier modules and given them the names of action potentials and local potentials. The neurons of the brain are constantly generating these electrical signals. These electrical signals can be detected by sensitive electrodes strategically placed on the skin of the scalp and recorded on an instrument known as an electroencephalograph.

11.4.1

Electroencephalogram

An electroencephalogram (EEG) is the tracing recorded by an electroencephalograph. Although the awake brain doesn't constantly produce a regular, repeating pattern like might be seen in an electrocardiogram, there are some measurable patterns that appear in different conditions, and the information in those patterns can be very useful in evaluating activity in the brain. Indeed, a flat EEG is a sign of clinical death. Based on the frequency (number of cycles per second) and amplitude (height of the waves) of the tracings, four typical EEG patterns can be discerned. If we were to record your EEG right now as you are reading this paragraph, we should detect beta waves. Beta waves have a frequency of 13-30 Hertz (cycles per second) and are common when the subject is awake with their eyes open and their brains engaged. Now close your eyes and try to relax and let your mind wander. Alpha waves would be recorded under these conditions. Alpha waves have a frequency of 8-13 Hz and a higher amplitude than beta waves. These are the only two types of waves that we should see in normal, healthy adults while awake. Theta waves (4-7 Hz) can be seen in the early stages of sleep and in young children. Finally, delta waves (0.5-4 Hz) are only seen during deep sleep or in adults with serious brain injury. As the frequency decreases, the amplitude of the waves gets larger. The larger amplitude is because the activity is more synchronized. The sleeping cortex is not receiving stimulatory input from the reticular system, but it is receiving stimulation from pacemaker areas like the thalamus, which can initiate rhythms in the cortex that can be perpetuated even after the pacemaker activity stops (kind of like how a crowd keeps chanting even after the music stops playing). What do these wave patterns mean? We still don't know, but the theta and delta waves may play a role in blocking sensory input and allowing the restorative functions of sleep.

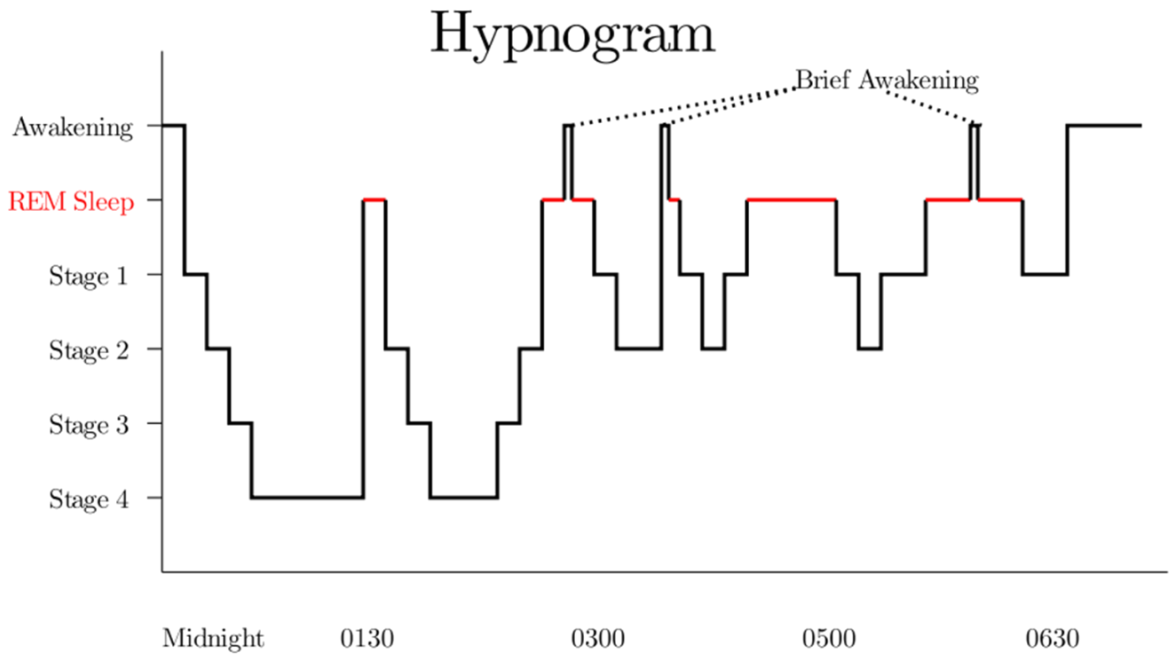
What is the significance of the alpha and beta waves? Consider the research from Harvard professor Eric Mazur in which he continually monitored brain patterns of students for one week. The data showed that brain waves were almost completely flat during two events, watching television and listening to a classroom lecture. Coupled with the Australian study of 11,000 adults that showed that for every hour of television watched 22 minutes of lifespan was lost, one begins to wonder if students will ever survive college as the two biggest time sinks are lecture and television. Studies like this don't capture what really happens in all situations though. We believe that BYU-Idaho students are closer to beta waves during their lectures ☺.

11.4.2

Sleep

One of the mysteries of science is why we have to sleep. Several theories have been proposed as to why we sleep. Some of the more probable include: 1) When one sleeps, they are not as likely to be seen and hunted by predators. 2) It may be a way of conserving energy. 3) It may allow the body to repair itself and/or recover from intense activities. 4) Sleep may allow the brain to work through emotional events and helps to eliminate accidental, repetitions and meaningless communications that have occurred throughout the day. 5) It may give the brain the opportunity to archive and store important memories and discard unnecessary ones. All of these are possible benefits and reasons for sleep, however, we really don't know exactly why we have to sleep. William Dement, a pioneer in sleep research and founder of Stanford's Sleep research center summed up why we sleep as follows, "As far as I know, the only reason we need to sleep that is really, really solid is because we get tired."

Let's examine what we know about sleep. There are two different kinds of sleep: rapid eye movement (REM) sleep and non-REM (NREM) sleep. In addition, NREM sleep can be divided into four stages (stages 1-4). In young healthy adults that sleep regularly (probably those who live alone or got married young) a typical night's sleep starts with NREM sleep, passing through all 4 stages and then quickly reverts to REM sleep. The stages of sleep will then alternate between REM and NREM throughout the night. The cycle of REM sleep occurs about every 90 minutes. Each time REM sleep is reached it lasts a little longer. For instance, the first REM stage lasts about 5 to 10 minutes and the final REM stage can last up to an hour.



Title: Sleep Hypnogram.svg; Author: RazerM; Site: https://commons.wikimedia.org/wiki/File:Sleep_Hypnogram.svg; License: GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation
The different stages of sleep are marked by changing patterns of the EEG. During stage 1, 2 and 3 of NREM (light sleep), the predominant brain waves are theta waves (4-7 Hz). In Stage 4 (Deep Sleep) delta waves (.5-4 Hz) predominate and vital signs are at their lowest. This is the stage where sleep walking and night terrors occur. The normal progression takes about 30-45 minutes to reach this stage of deep sleep.

Upon reaching stage 4 sleep we cycle back to enter the phase of REM sleep. It is during this phase that we have our most vivid dreams. Also during REM sleep skeletal muscles are paralyzed (probably so that we don't act out our dreams) except for eye movements. Heart rate and respiration also increase and the brain uses even more oxygen than during the awake state.

So, do we really need to sleep? Well, individuals that do not get REM sleep become moody, depressed and even exhibit various types of personality disorders. (Does that sound like one of your roommates?) Some sleep medications increase NREM sleep but can actually decrease REM sleep, worsening their condition and potentially increasing their risk for developing other health problems. I like what Freud says about dreaming – that we are able to act out behaviors we wouldn't do while awake and to work through and resolve anxieties and fears.

And how much sleep do we need? Although we don't have a definitive answer, we do know that sleep requirements change with age and other factors. For example, infants require 16 hours while adults need 7-9 hours.

Despite the fact that we still don't understand the exact role of sleep there have been numerous scientific studies focused on sleep that show a correlation between inadequate sleep and various health conditions, including:

- Alzheimer's Disease
- Depression
- Anxiety
- Heart Disease
- Hormone Imbalances
- Immune Suppression
- Cancer
- Glucose Intolerance
- Many others

In 2007 Jeffrey R Holland said, "No misfortune is so bad that whining about it won't make it worse." (April 2007 General Conference) Sleep seems to be so important that we could modify that statement to state that no illness is so bad that poor sleep won't make it worse! Future studies will help us understand the relationship between sleep and wellness but the evidence certainly shows that sleep is one of the most important pillars of health! This is troubling in light of data showing that in 1942 the average American adult slept for 7.5 hours compared to 6.5 hours today. To wake up, neurons in the reticular formation begin to release chemicals that begin to arouse the cortex. As stated earlier, caffeine stimulates the reticular formation giving a jump start to the system. It does this by blocking adenosine receptors and thus preventing drowsiness. As a result, 68 million people in the US have a cup of coffee within one hour of waking up. The alternative is to sleep consistently so that the body develops its own natural "wake up" cycle.

11.4.3

Memory and Learning

A unique characteristic of humans and some animals is the ability to alter behavior based on past experiences. Arguably, some humans are better than others at this behavior and some seem to never learn, but by definition, **learning** is the acquisition of information and **memory** is the retention and storage of that information. Generally, memory is categorized as either **short-term** or **long-term memory**. Short-term, as the name implies does not remain with us long. It is usually limited to only a few bits of information, generally 7-12. It may only be retained for a few seconds to a few minutes. For example, you can remember a telephone number long enough to dial it but if the line is busy and you want to try again in a few minutes you usually have to look it up again. A special form of short-term memory is **working memory**. Working memory allows us to correlate our current situation with our experiences so that we can function. For example, if you need to cross the street you look to your left and see that no cars are coming, you then look to the right and see no cars are coming so you proceed to cross. Without functional working memory you would not remember what you saw when you looked to the left, making it very difficult to get across the street.

Long-term memory is information that is retained for extended periods of time, up to a lifetime. Unlike short-term memory, there seems to be no limit how much information we can store in long-term memories. The hippocampus has been shown to be essential for generating long-term memories. Indeed, in individuals who have had part or all of their hippocampus removed to treat certain types of epilepsy, generation of long-term memory is very difficult. They can remember things learned before the surgery but have a difficult time generating new long-term memories after the surgery. The creation of long-term memories appears to involve physical changes in the neurons such as the formation of new synapses or reinforcing existing synapses creating a permanent **memory trace**. Long-term memories seem to be stored in areas of the brain that are most easily accessible by the regions that need them. For example, visual memories are stored in the occipital cortex and memories of sound are stored near the temporal cortex. The process of converting short-term memories to long-term memories is known as **consolidation**. Students often ask what the best techniques for consolidating memories are. The answer is simple, though no one we really wants to hear it. The answer is repetition. The best way to create long-term memories is to review the material over and over.

Long-term memories can be divided into two categories: **explicit** or **declarative memory** and **implicit** or **procedural memory**. Explicit memories are things like what you are trying to learn in school, names, dates, processes, etc. Explicit memory involves essentially those things that can be expressed verbally. Recalling explicit memories requires conscious processing and recalling of the memory. The hippocampus is intimately involved in generating explicit memories. Among the different types of long term memory, declarative (explicit) memories are most easily lost over time.

Implicit memory is memories of things that are not expressed verbally and that we do not have to consciously recall. For example, once we learn how to ride a bike or even walk we do not have to think about it anymore, it is automatic (even reflexive). Some memories start as explicit memory, like when you were first learning to ride a bike, but then become implicit once we have mastered the task. Once again repetition and practice are the best way to generate implicit memories. The amygdala and cerebellum have been shown to play important roles in generating implicit memories.

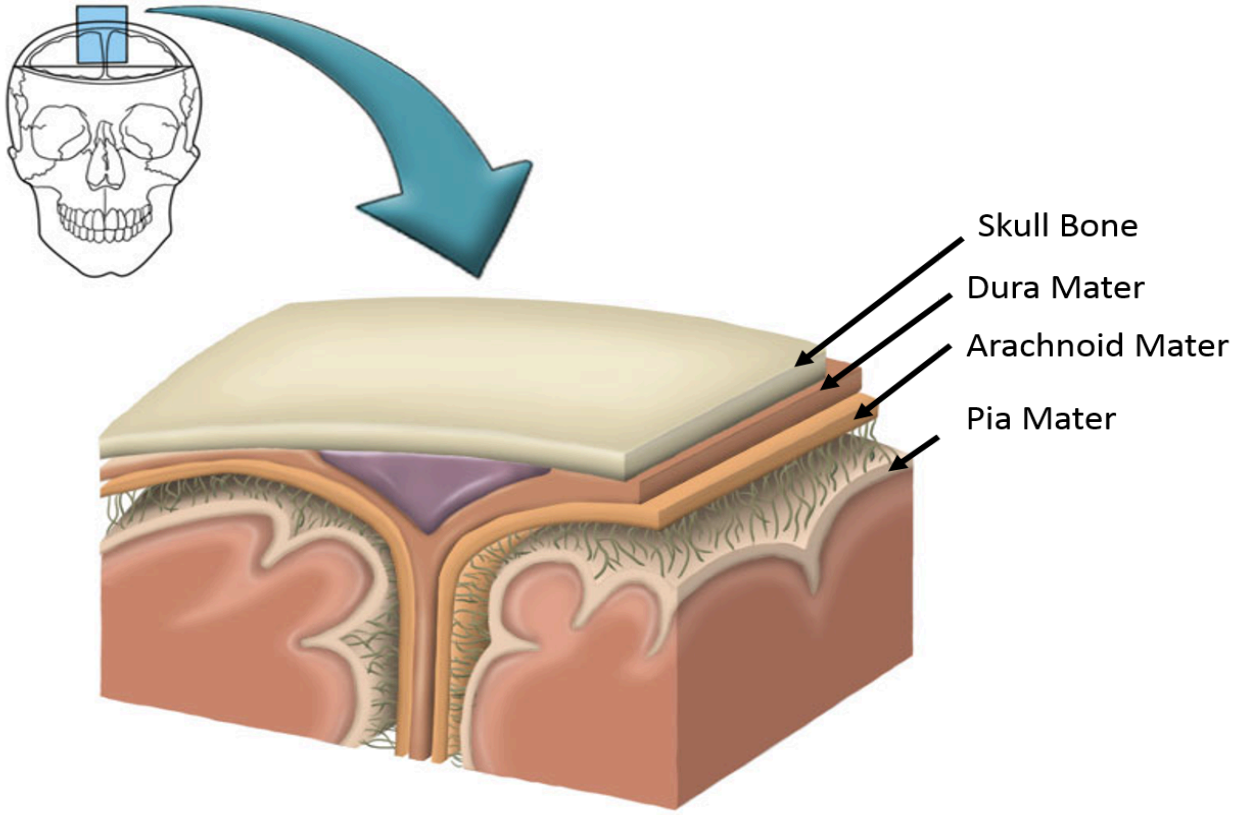
11.5

THE MENINGES, CEREBRAL SPINAL FLUID AND CRANIAL NERVES

11.5.1

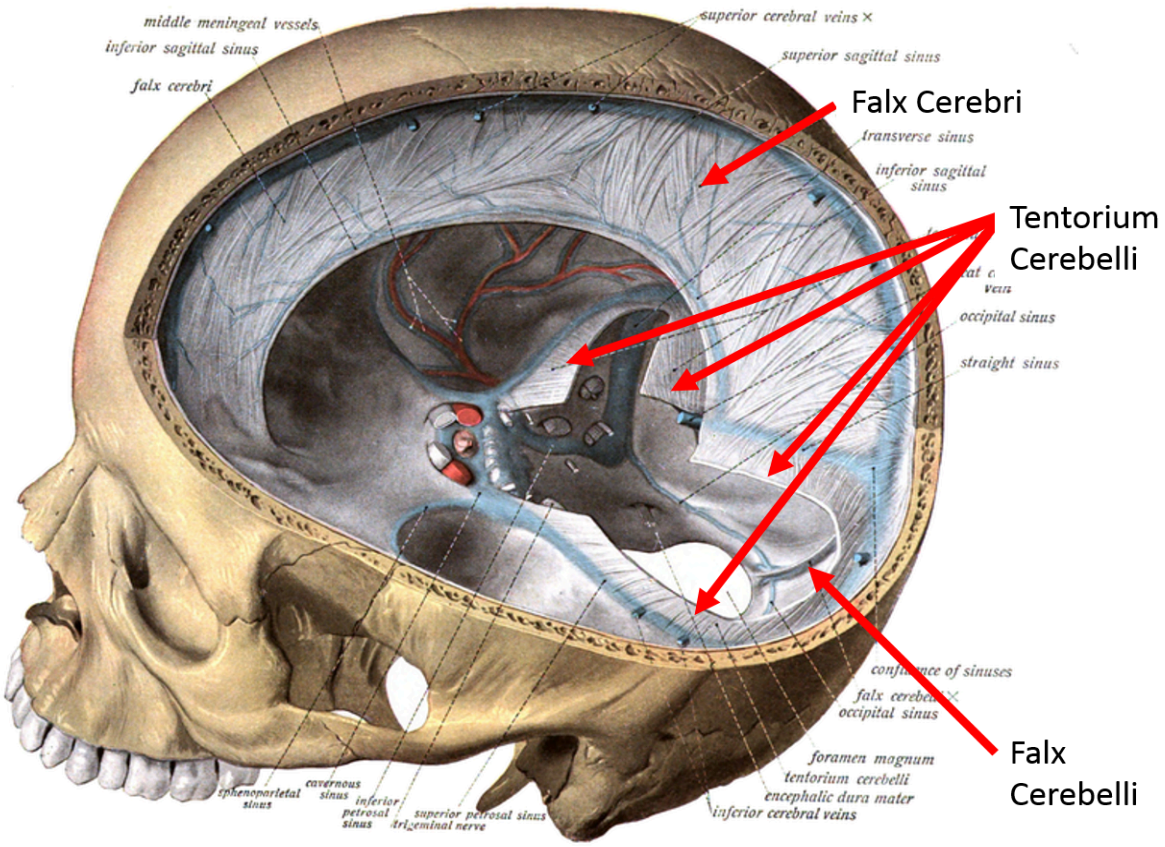
The Meninges

The body goes to great lengths to feed and protect the central nervous system. Part of the protective mechanism is found in the meninges which are membranes that envelop the central nervous system. The meninges consist of three layers: the **dura mater**, the **arachnoid mater** (**membrane**), and the **pia mater**.



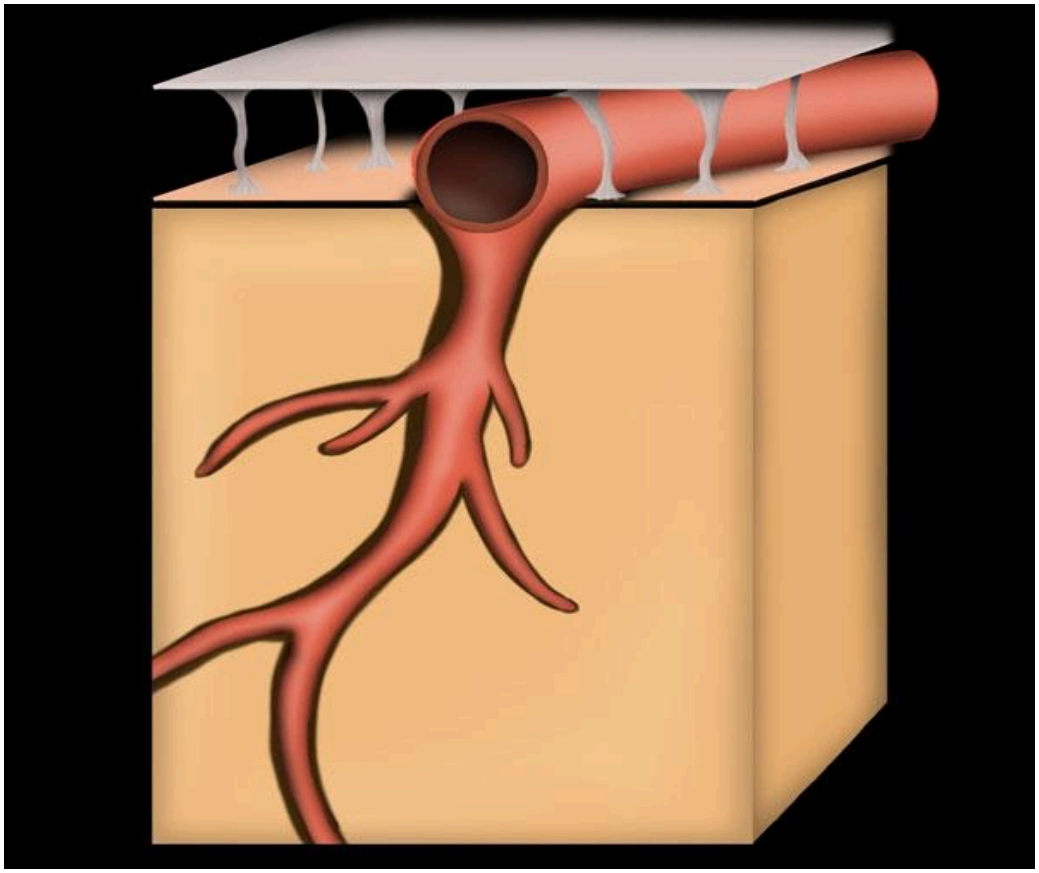
Layers of the Meninges. *Image was drawn by a BYU-Idaho student Winter 2014.*

The dura mater is a thick, tough, and durable membrane composed of dense fibrous connective tissue. The dura mater of the brain is composed of two layers, an outer **periosteal layer** that connects to the inside of the skull, and a deep **meningeal layer.** In certain areas these two layers separate and the meningeal layer forms folds that extend into the brain forming physical partitions in the brain. The **Falx cerebri** separates the two cerebral hemispheres, the tentorium cerebelli separates the occipital lobes of the cerebrum from the cerebellum and the falx cerebelli separates the two cerebellar hemispheres.



Falx Cerebri, Tentorium Cerebelli, Falx Cerebelli. *From Sobotta's Human Anatomy 1908. Public Domain.*

The arachnoid mater or arachnoid membrane is directly under the dura mater. These two membranes are not physically connected and there is a space (more like a virtual space) between the dura mater and the arachnoid mater. The arachnoid mater is a very thin and transparent membrane that lies on top of a fluid filled space directly inferior to the membrane. This space, the **subarachnoid space**, is filled with **cerebrospinal fluid**. The arachnoid membrane together with the cerebral spinal fluid helps to cushion the central nervous system and fits like a loose sac over it.



Layers of Brain: Dura Mater, Arachnoid Mater, Subarachnoid Space, Pia Mater, Brain.
Image drawn by BYU-Idaho Student, Spring 2014.

The arachnoid mater got its name from the many processes that extend down from the membrane through the subarachnoid space to the pia mater on the brain surface. These processes are very fine and look a bit like spider web fibers.

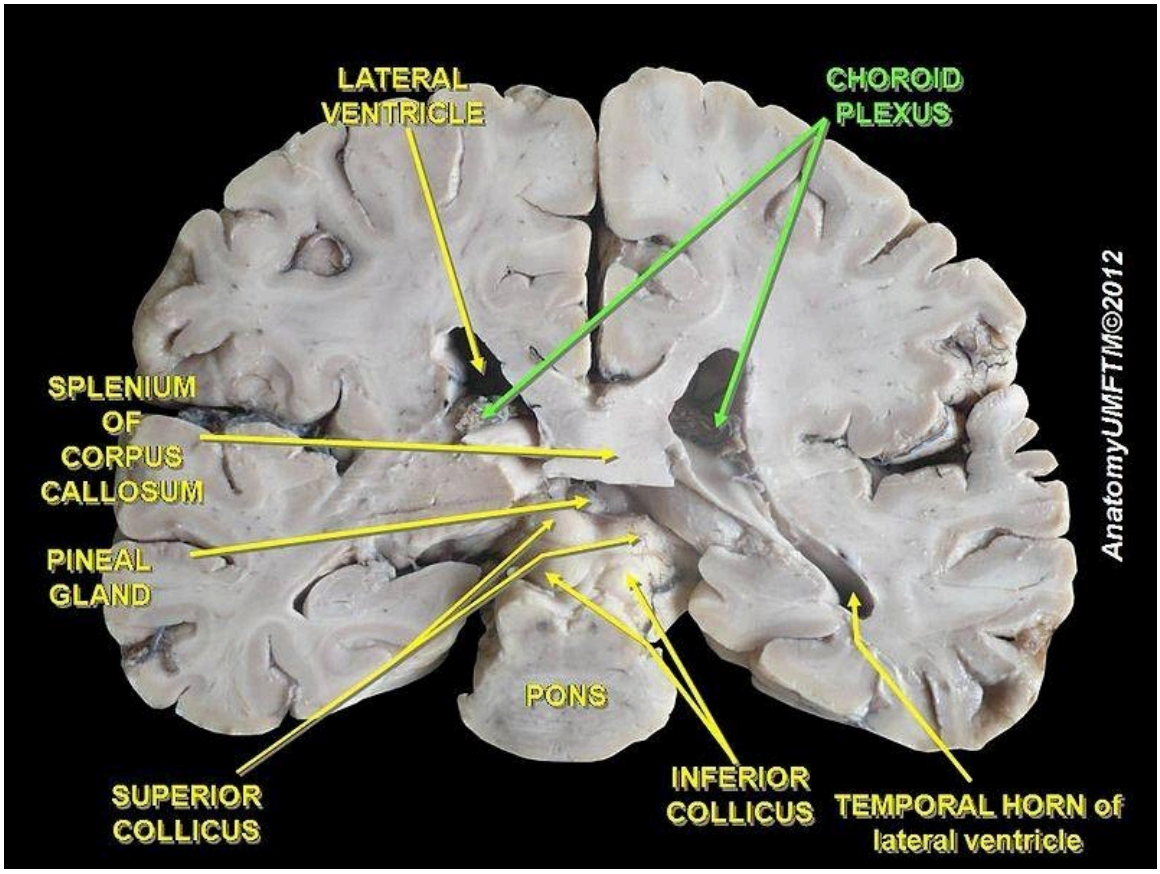
The pia mater is a very delicate membrane that adheres to the surface of the brain and spinal cord, following the contours (gyri and sulci) of the brain. The cerebral spinal fluid of the central nervous system sits on top of the pia mater (and underneath the arachnoid membrane).

There are some subtle differences between the meninges of the brain and the spinal cord, primarily with the dura mater. First, the dura mater of the spinal cord is composed of just a single layer, rather than two like we described in the brain. Second, the dura mater does not connect to the bones of the vertebra, instead, there is a space between the vertebra and the dura mater called the **epidural space**. This space is filled with adipose that acts as a cushion and helps protect the spinal cord.

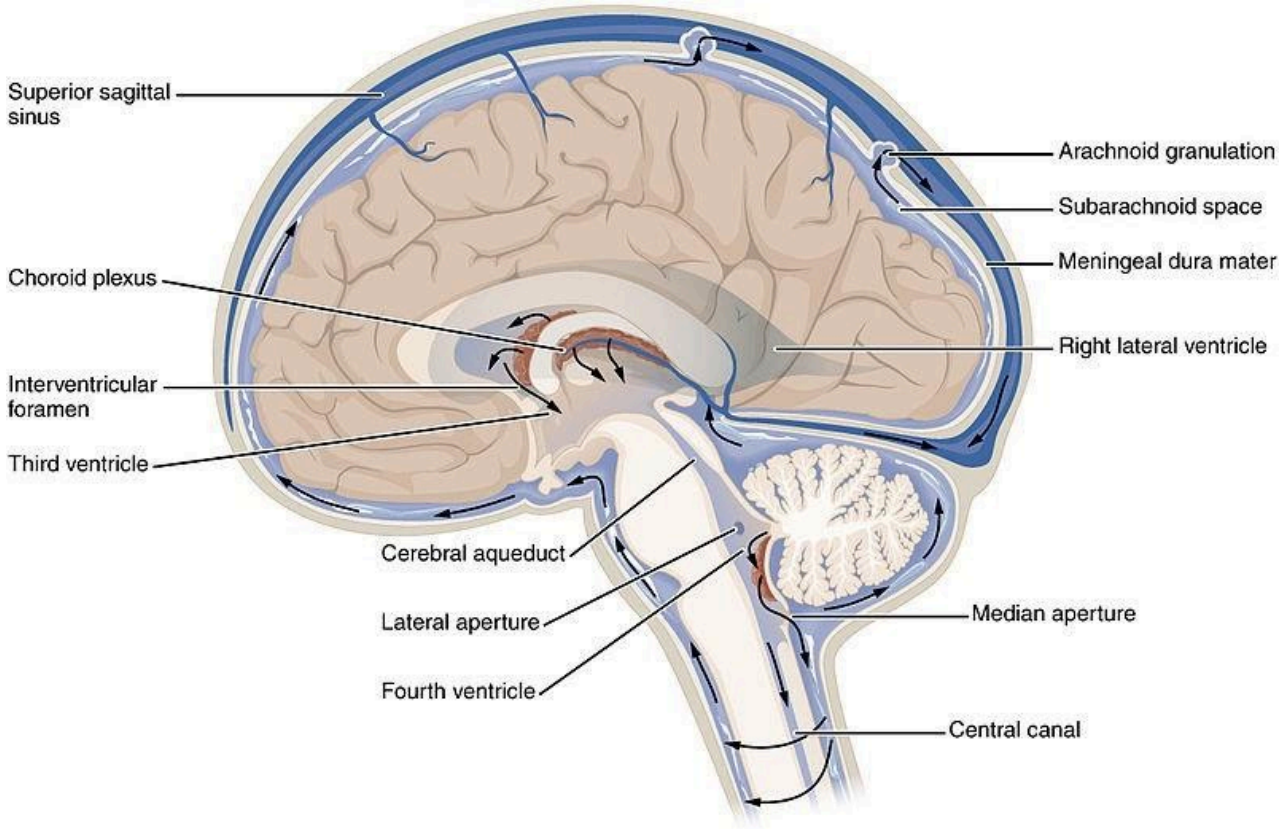
11.5.2

Cerebrospinal Fluid

Cerebrospinal fluid (CSF) is the “blood” of the brain and spinal cord. It is produced in a structure called the choroid plexuses by ependymal cells.



Author: Anatomist90; Wikimedia Commons download Spring 2014; License: Creative Commons Attribution-Share Alike 3.0



Brain Anatomy. *Author: OpenStax College; License: Creative Commons Attribution 3.0 Unported license.*

These structures are found lining the ventricles of the brain. The majority of CSF is produced within the two lateral ventricles. The fluid moves from the lateral ventricles to the third ventricle which is found in the diencephalon. The fluid then moves down inferiorly through a relatively narrow canal called the cerebral aqueduct. After exiting the cerebral aqueduct, the CSF fluid is in the 4th ventricle. From the 4th ventricle, some CSF fluid continues inferiorly to fill the central canal of the spinal cord. The rest of the CSF fluid exits the 4th ventricle through the apertures (median and lateral). After exiting through these apertures, the CSF is now in the arachnoic space where it can travel all the way around the brain and spinal cord. This arachnoid space is found between the arachnoid mater and the pia mater. The total amount of CSF ranges from 100-600ml, but it is produced and reabsorbed at a rate of 500ml per day. Ependymal cells produce CSF from the blood plasma making the contents of the CSF almost identical to the blood plasma, with the exception of proteins (very little if any proteins found in normal CSF).

CSF is drained into a structure called the **Superior Sagittal Sinus** through small granulations called “Arachnoid Granulations”. The superior sagittal sinus drains the CSF with venus blood from the brain down and out of the skull via the jugular vein.

The function of the CSF can be summarized into five categories:

1. **Buoyancy:** The CSF helps to suspend the brain which reduces the effective weight by 95%. This allows the brain to exist in a state called neutral buoyancy. In less scientific terms, it allows the weight of the brain to not squish itself.

2. **Electrolyte and circulatory balance:** The CSF is very important in maintaining the homeostatic balance of electrolytes and glucose.

3. **Protection:** The properties of water allow the CSF to serve as a protection against jolts.

4. **Circulation:** Although the CSF is a very low-pressure system, it is an effective method of moving nutrients and waste around. It also helps to facilitate blood perfusion.

5. **Waste removal:** The CSF has proven essential in flushing metabolic waste. Recently researchers have shown that the flushing of waste is increased during sleep.

Sometimes a breach can occur, such as physical trauma or a lumbar puncture resulting in leakage of the CSF. This can change the pressure and allow the brain to not be as buoyant, which squishes on nerves and causes a variety of symptoms. In addition, in other cases a condition called hydrocephalus can occur. Hydrocephalus is the accumulation of CSF because of impaired flow or excessive production of CSF. This can also result in pressure changes and if it occurs in the fetus it can result in an enlarged head. Hydrocephalus in an adult must be immediately corrected.

11.5.3

Traumatic Brain Injury and Cranial Bleeds

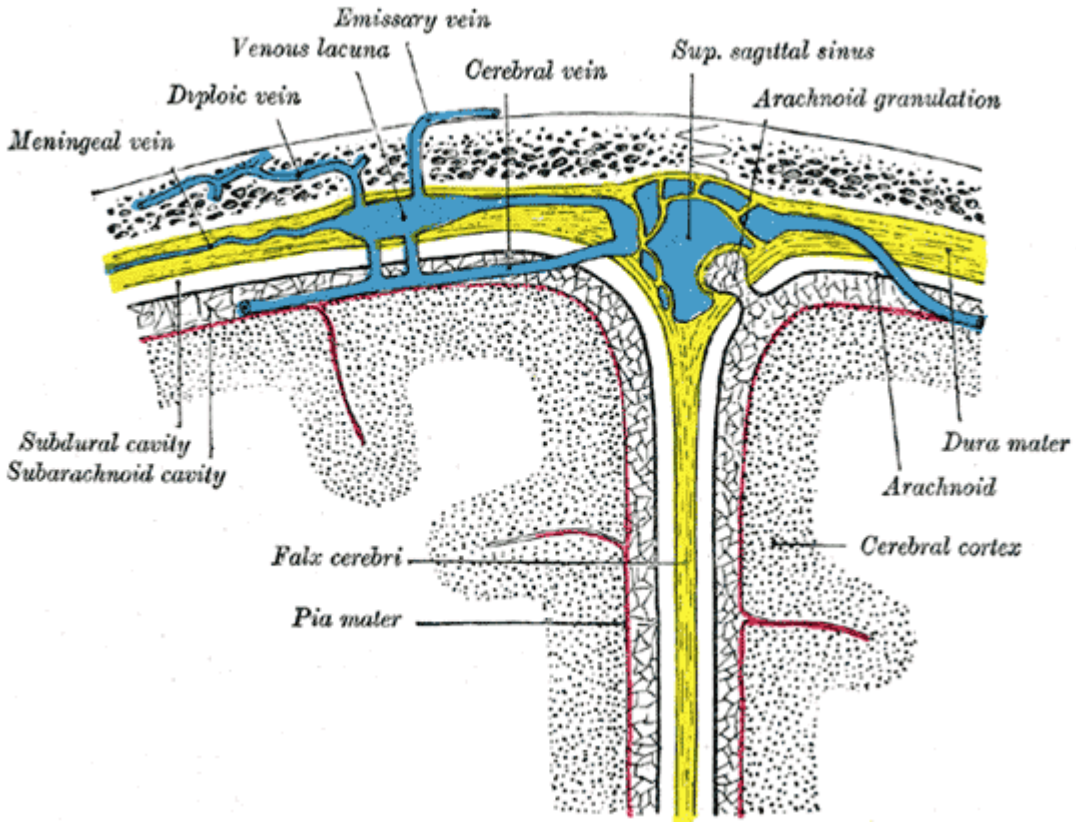
There are different types of hemorrhages that can occur in the central nervous system. We will discuss the four major types: epidural, subdural, subarachnoid and intracerebral hemorrhages.

Epidural Hemorrhages

Epidural hemorrhages occur between the dura mater and the skull and are usually very rapid because the hemorrhage comes from damage to the arteries along the inside of the skull. Recall looking at the skulls in lab and seeing the grooves where the arteries of the skull once ran. These arteries are under high-pressure vessels bleed rapidly when damaged. This bleeding results in a hematoma, which strips the dura membrane off the skull as it expands, causing intense headaches. More seriously is the compression of the nervous system as the hematoma expands against the skull. The most common cause of epidural bleeds is a skull fracture which lacerates these arteries. Epidural hemorrhages can be fatal if left untreated. Treatment is done by surgically draining or removing the hematoma to relieve pressure on the brain.

Subdural Hemorrhages

Subdural hemorrhages occur between the dura mater and the brain. These result from tears in the veins that cross the subdural space in response to a head injury, especially rotational or linear forces. Subdural hemorrhages are classic injuries found in shaken baby syndrome and severe whiplash. They are also more common in people taking aspirin since aspirin inhibits blood clotting. They can be subdivided into acute, subacute and chronic subdural hematomas depending on the severity of the hemorrhage. Acute hematomas develop rapidly and are the most severe with a mortality rate of 60 to 80%. Subacute hematomas fall into the same category although slightly less severe. Chronic subdural hematomas develop over a period of days to weeks and often result from minor head trauma (like a concussion). Symptoms of subdural hemorrhage typically have a slower onset than epidural bleeds because the bleeding comes from veins instead of arteries.



Layers and Veins Structure of the Cerebral Cortex. *Grays Anatomy: Public Domain*

To understand subdural hemorrhages better, it helps to examine the anatomy. Notice that veins from the brain cross the subarachnoid space and pierce both the arachnoid and dura mater to finally dump venous blood in the superior sagittal sinus. When whiplash, shaking or any type of rapid, intense linear motion occurs, the brain moves inside the skull. This type of movement will put tension on the veins where they pierce the arachnoid and dura mater membranes. The arachnoid membrane moves relatively easy compared to the dura mater. A shearing effect can occur where the vein is broken off just superior to the arachnoid membrane. Blood slowly starts to fill the space between the arachnoid mater and the dura mater.

Subarachnoid Hemorrhages

Subarachnoid hemorrhages occur in the area between the arachnoid membrane and the pia mater that surrounds the brain (remember this is where the cerebral spinal fluid is found). There are many arteries in the subarachnoid space as it is this space that blood vessels like the internal carotid arteries enter. The bleeding may occur by spontaneous rupture or as a result of head injury. The most common symptom is called the thunderclap headache, or one that develops immediately within seconds and feels like a kick in the head.

Intracerebral Hemorrhages

Intracerebral hemorrhages occur within the brain tissue itself and usually involve very small blood vessels. This type of hemorrhage may be caused by trauma or spontaneous rupture. Symptoms are associated with the functional area of the brain that is experiencing the trauma. Intracerebral hemorrhages are the second most common cause of stroke and the risk of experiencing this type of hemorrhage is increased by high blood pressure and diabetes.

11.5.4

Cranial Nerves

There are lots of great pictures of cranial nerves in any internet search. CLICK HERE to load an image search of cranial nerves and check some of the thumbnails out to help you get an idea of where these nerves are at. You will also study the location and anatomy of these nerves in Bio 264 lab.

Remember the friend that you took to the emergency room after she got hurt sledding? When the doctor shined a light into her eyes to check her reflexes he was actually performing a small part of a cranial nerve examination. Doing an examination of the cranial nerves can provide valuable clinical information about the state and condition of the nervous system. There are twelve pairs of cranial nerves that originate in the brain and carry information to and from the brain. The cranial nerves are designated by Roman numerals (I - XII) and by names. The numbering starts with those most superior and anterior and progresses posteriorly and inferiorly. The names usually correspond to either the function or the structure of the nerve, hence, the Optic nerve is involved with vision and the Trigeminal nerve has 3 branches. These nerves can carry both sensory and motor information, just as we have seen with spinal nerves. We can subdivide the sensory information further into 1) special senses and 2) general senses. The motor information can also be subdivided into 1) somatic motor and 2) parasympathetic. In contrast to the spinal nerves, however, not all cranial nerves carry both sensory and motor information. In fact, three cranial nerves carry purely sensory information and four cranial nerves carry almost entirely somatic motor information. The remaining cranial nerves carry some combination of sensory, somatic motor and parasympathetic information. Because of the clinical importance of the cranial nerves we will discuss further the major functions of each one and some of the common symptoms observed when the nerves are damaged.

Cranial Nerve 1 (CN I): Olfactory

Major Function: Sensory - Smell (olfaction)

Lesion: Loss of smell on the affected side

*Note: Loss of smell doesn't necessarily confirm a CN I lesion as an upper respiratory tract infection etc. could also decrease olfaction.

Cranial Nerve 2 (CN II): Optic

Major Function: Sensory - Vision

Lesion: Blindness on affected side and loss of pupillary light reflex (Described Later)

Cranial Nerve 3 (CN III): Oculomotor

Major Function: Somatic Motor to four of the six extrinsic muscles that move the eye

ANS Innervation: Parasympathetic to sphincter pupillae muscle for constriction of the pupil

Lesion: Eye deviation causing double vision, pupil dilation and loss of pupillary light reflex

Cranial Nerve 4 (CN IV): Trochlear

Major Function: Somatic Motor to superior oblique eye muscle

Lesion: Eye deviation causing double vision

Cranial Nerve 5 (CN V): Trigeminal

Major Function: Sensory - General sense from the face and forehead (including sensation of much of the mouth and anterior 2/3 of the tongue)

Motor: Somatic Motor to muscles of mastication (chewing muscles)

Lesion: Loss of sensation in face and forehead or increased sensitivity to pain known as Trigeminal neuralgia. (Described Later). Also, muscle weakness of the muscles of mastication

Cranial Nerve 6 (CN VI): Abducens

Major Function: Somatic Motor to lateral rectus eye muscle

Lesion: Medial deviation of the eye causing double vision

Cranial Nerve 7 (CN VII): Facial

Major Function: Sensory - Taste from the anterior 2/3 of the tongue and Motor - Somatic Motor to the muscles of facial expression

ANS Innervation: Parasympathetic to salivary glands and lacrimal glands

Lesion: Facial paralysis often called Facial or Bell's palsy. (Described Later). Decreased ability to taste (particularly on the anterior 2/3 of the tongue). Decreased salivation and lacrimation (tearing)

Cranial Nerve 8 (CN VIII): Vestibulocochlear

This nerve is composed of fibers from two branches: the vestibular nerve and the cochlear nerve, each with specific functions

Major Function: Sensory - Vestibular branch senses balance. Cochlear branch if for hearing.

Lesion: Vestibular-If only the vestibular branch is damaged it would result in loss of balance and dizziness (vertigo). Cochlear-If only the cochlear branch is damaged it would result in loss of hearing. If the lesion occurs after the two branches converge then you could have a combination of the above symptoms.

Cranial Nerve 9 (CN IX): Glossopharyngeal

Major Function: Somatic Motor to swallowing muscles of the throat and Sensory - Taste to the posterior 1/3 of the tongue and Sensory from the pharynx, carotid body and carotid sinus

ANS Innervation: Parasympathetic to salivary glands

Lesion: Trouble swallowing, loss of taste (particularly to posterior 1/3 of tongue), decreased ability to sense and respond to blood pressure changes and decreased salivation

Cranial Nerve 10 (CN X): Vagus

Major Function: Motor - Somatic Motor to throat muscles involved in swallowing and speech and Sensory - Taste from the posterior tongue. Also, sensory from throat, thoracic and abdominal organs

ANS Innervation: Parasympathetic to thoracic and abdominal organs regulating things such as heart and respiratory rate and gastrointestinal peristalsis etc.

Lesion: Trouble swallowing and hoarse speech uvula deviation away from side of lesion

Cranial Nerve 11 (CN XI): Accessory

Major Function: Somatic motor to sternocleidomastoid and trapezius muscles

Lesion: Muscle weakness and trouble turning the head and elevating the scapula

Cranial Nerve 12 (CN XII): Hypoglossal

Major Function: Motor - Somatic Motor to tongue and throat muscles

Lesion: Tongue deviation toward the side of the lesion; trouble manipulating food with tongue and trouble swallowing.

Some Important Cranial Nerve Functions and Clinical Conditions

Pupillary Light Reflex

The pupil diameter is closely regulated and responds to the amount of light available. The pupil will dilate in a dark environment to allow in more light and constrict in a light environment to restrict the amount of light entering the eye. This dynamic control has two branches. The afferent (sensory) limb of the reflex is regulated via CN II, which sends action potentials to the control center in the midbrain regarding light intensity. The midbrain then sends signals through the efferent (motor) limb of the reflex, which is CN III, to constrict the pupil. Dilation of the pupil is achieved via a sympathetic nerve which exits the CNS in the spinal chord and is not mediated by a cranial nerve. **Clinical Manifestation:** When you shine a light into a patient's left eye the optic nerve should increase signals to the midbrain which will then cause the oculomotor nerve to stimulate the constrictor pupillae muscle to contract, thus constricting the pupil of the left eye. This is referred to as the direct light reflex. In addition to pupil constriction in the left eye, the pupil of the right eye will also constrict, which is known as the consensual light reflex. Any deviation from this pattern represents a pathological condition that would warrant further investigation.

Trigeminal Neuralgia (Tic Douloreux)

CN V gets its name from the fact that it has three branches, trigeminal means "three twins." The three branches (V1, V2, and V3) are responsible for providing innervation to specific regions of the anterior head. V1 (ophthalmic) supplies sensory innervation to the forehead down to the nose. V2 (maxillary) supplies sensory innervation to the maxillary region inferior to the nose and superior to the lower jaw. V3 (mandibular) supplies sensory innervation to the mandibular region or the lower jaw and anterior to the ear, as well as somatic motor innervation to the muscles involved in chewing. The maxillary and mandibular branches of the trigeminal nerve are also responsible for supplying sensory innervation to the teeth.

When you "visit" the dentist to fill a cavity the dentist tries to minimize the trauma by anesthetizing the area that he is going to subsequently abuse with various power tools! The most common way this is done is by performing a "nerve block" in which inject lidocaine, or some other local anesthetic is injected into the area surrounding either the maxillary nerve, if they need to work on your upper teeth, or the mandibular nerve, if they are working on your lower teeth. This "nerve block" literally blocks or prevents action potentials from being sent to your brain.

Trigeminal neuralgia involves intense episodic pain in any or all three facial areas supplied by CN V. It has been described as one of the most intensely painful conditions known to man! It is characterized by hypersensitivity of the nerve to the point where a light touch on the face or even a mild breeze can cause intense, debilitating pain. Those with

trigeminal neuralgia often describe the pain as burning, electrical, stabbing, crushing, or even exploding pain. And if that doesn't sound bad enough it is also quite difficult to control. Fortunately, there are several pharmacologic treatments to help manage the pain.

The leading theory to explain the cause of this condition involves compression of the myelin sheath around the nerve, likely due to an enlarged artery or an aneurysm. This compression can lead to the destruction of the myelin sheath, causing the nerve to become hypersensitive to the slightest stimulation. Likewise, this can also make it difficult for the trigeminal nerve to stop the afferent pain signals once they have begun.

Facial Palsy (Bell's Palsy):

Bell's palsy is a condition that results in partial or complete facial paralysis on one side of the face, although some cases can manifest bilaterally. This is a result of a lesion of the Facial Nerve (CN VII), most likely due to inflammation although the precise cause is unknown. The leading explanation is that a dormant herpes viral infection becomes reactivated causing the facial nerve inflammation. Stress, trauma, environmental and other factors may precipitate the reactivation of the virus. There are other causes of facial paralysis, the most common being a brain tumor or stroke.

Treatment: The symptoms of Bell's palsy generally resolve on their own over the course of days to weeks, but corticosteroids administered early after onset of symptoms has been shown to improve recovery. Because the affected eyelid is often unable to close, it is important to prevent drying of the cornea by using eye drops and/or physically closing the eye! (Taping the eye closed and using an eye patch can be helpful and who doesn't love a pirate, arrrrrr!)

Speaking of pirates, it has been rumored that pirates wore eye patches so that as they went from the deck where it was very bright to the rooms below deck where it was quite dim they could remove the patch and maintain their ability to see. Based on the consensual light reflex described above, it is unlikely that pupil constriction/dilation would be the mechanism allowing pirates to see as they move from light to dark environments. Keep this in mind as you learn about vision in subsequent modules and, in particular, the rhodopsin cycle.

A cranial nerve examination is an important part of any physical examination where you suspect that there might be some level of brain trauma, but particularly when there are no clear symptoms that suggest brain injury. Because of their connections with various parts of the brainstem and cerebrum, functional deficits related to one of the cranial nerves can provide valuable insight into the location and severity of damage. Often a medical provider can perform a brief cranial nerve exam at the same time that they are performing their routine physical exam. For example, when they look at your throat and you say "ahhhhhh" they are not only looking for redness and inflammation etc. associated with viral or bacterial infection. After all, do you think your throat would immediately become redder by doing this? The answer is no. The physician is really looking for elevation of the palate or deviation of the uvula, which would signal a cranial nerve problem. The following is a link to a YouTube video of a brief cranial nerve examination. https://books.byui.edu/-PfY (Video Transcriptions Available)

12.0

MODULE 12: SPECIAL SENSES

12.1

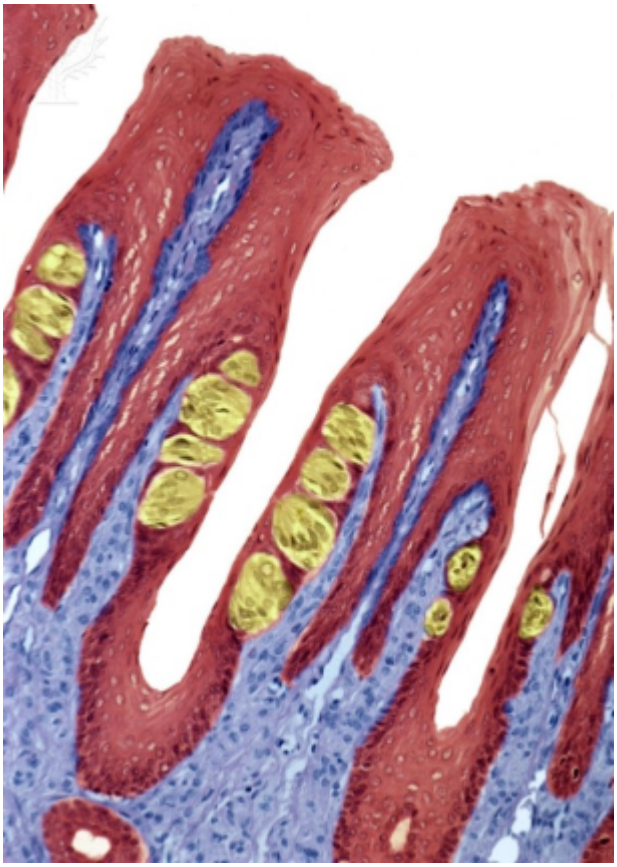
THE SENSE OF TASTE AND SMELL

12.1.1

Taste

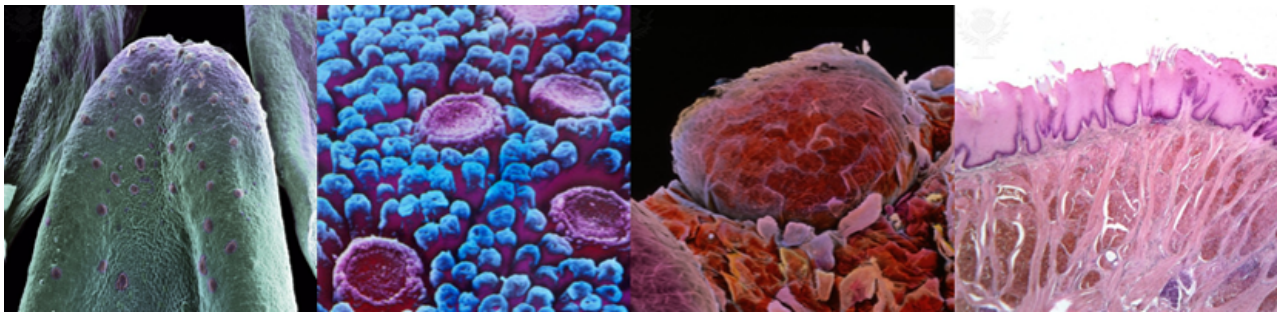
Have you ever sat down and eaten something that was just bliss? Perhaps a moist chocolate cake topped with chocolate cream cheese frosting and a fudge filled center, or a marinated chicken breast grilled with roasted red peppers, onions, Portobello mushrooms and Dijon sauce. Some foods have the ability to flood our system with all kinds of different tastes and textures, yet we can sometimes still distinguish the individual contributions of the specific ingredients. This myriad of sensation should not be fully credited to taste because the sense of olfaction and the ability to detect texture are also intimately linked to this perception. In fact, you may have noticed that your mouth is starting to water as you contemplate a plan to obtain the previously mentioned food. The senses of our body-vision, taste, smell, hearing, touch, and equilibrium-are conduits through which we "feel" our world. Each sensory system is uniquely designed with receptors that detect various environmental stimuli and then convert those stimuli to action potentials. That's right; all stimuli that the body detects are ultimately converted to action potentials. In response to the action potential, the brain can interpret and allow us to feel the stimulus. In this module, we will discuss the special senses of taste, smell, vision, hearing and equilibrium. The special senses are those in which their receptors are localized in a specific, fairly small area of the body. In contrast to the special senses, the general senses have receptors that are widespread. The general senses include the sensations of touch, pressure, temperature, pain, and proprioception (movement and position of the body).

Like olfaction, taste (gustation) is a chemical sense responding to different compounds in the foods we eat. It is estimated that humans can distinguish 4000 to 10,000 different chemicals with only five different taste modalities: salt, sweet, bitter, sour, and umami (Japanese for delicious). The receptors responsible for transmitting these taste qualities are located mainly on the dorsal surface of the tongue within structures called **taste buds** which are housed in epithelial projections called **lingual papillae**.



Microscopic View of Tongue. *© 2013 Encyclopædia Britannica, Inc. Downloaded from Image Quest Britannica; BYU-Idaho.*
The figure above is an enhanced color light micrograph of a thin section through the tongue. The yellow colored structures are taste buds.

Four types of lingual papillae are observed on the human tongue: (1) **circumvallate papillae** (large groove surrounding the papilla), (2) **fungiform papillae** (mushroom shaped), (3) **foliate papillae** (leaf shaped), and (4) filiform papillae (string shaped) so named based on the shapes of each (see image below). Only the circumvallate, fungiform, and foliate papillae contain taste buds; the filiform provide a rough surface for moving food around in the mouth. On average, people have 2000 to 5000 taste buds, but in some exceptional cases, 20,000 have been observed. This has prompted some to propose taste classifications for people that include "non-tasters," "tasters," and "super tasters." Taste, however, is not an all or none phenomenon and the classifications of "non-taster" and "super tasters" simply represent populations of people on opposite ends of the spectrum. Interestingly, some "supertasters" taste and review food for a living and have insured their tasting abilities (job security) for astronomical amounts, like 1 million dollars.



The Tongue. *© 2013 Encyclopædia Britannica, Inc. Downloaded from Image Quest Britannica; BYU-Idaho.*

Moving from left to right, the first figure is an artificially colored scanning electron micrograph (SEM) of a tongue (rat). The second figure is a SEM of the tongue illustrating circumvallate papillae, the third figure is a SEM of fungiform papillae and the fourth figure is a light micrograph of filiform papillae.

Each taste bud contains 50-150 **receptor cells** (taste or gustatory cells) sandwiched in between supporting cells and basal cells. Receptor cells are specialized epithelial cells that synapse with neurons. They have the ability to excite their associated neuron when they interact with a taste stimulus, generating action potentials in the neurons. The receptor cells have a relatively short life span and must be replaced about every 10 days. The role of the basal cells in the taste buds is to replace the worn out receptor cells. The taste stimulants (**tastants**) that react with receptor cells most efficiently are sodium chloride (salty), sucrose (sweet), hydrochloric acid (sour), quinine (bitter) and monosodium glutamate (umami). There is a unique receptor for each of these taste modalities. However, other molecules can also interact with these receptors. For example, artificial sweeteners (e.g., saccharin and aspartame) are known to be 10,000 to 100,000 times more effective than sucrose at stimulating the sweet receptors. Each receptor cell is oriented in the taste bud such that the tip of the cell protrudes into the oral cavity through an opening called a taste pore. This orientation allows the apical end to respond to the tastant and the basal end to transmit the signal to the brain.

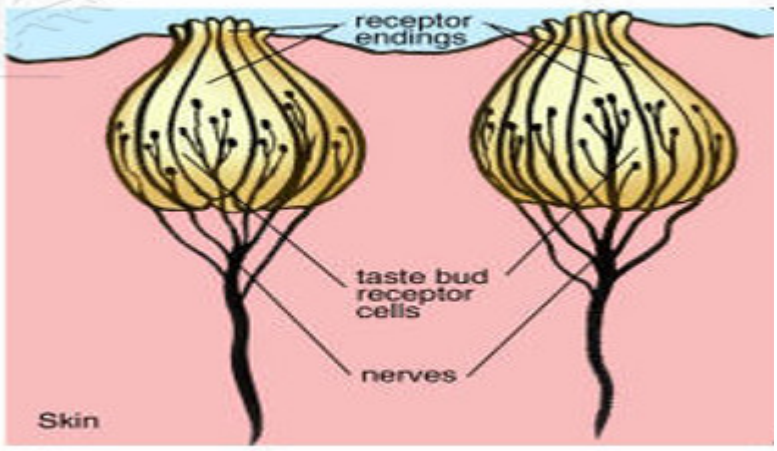
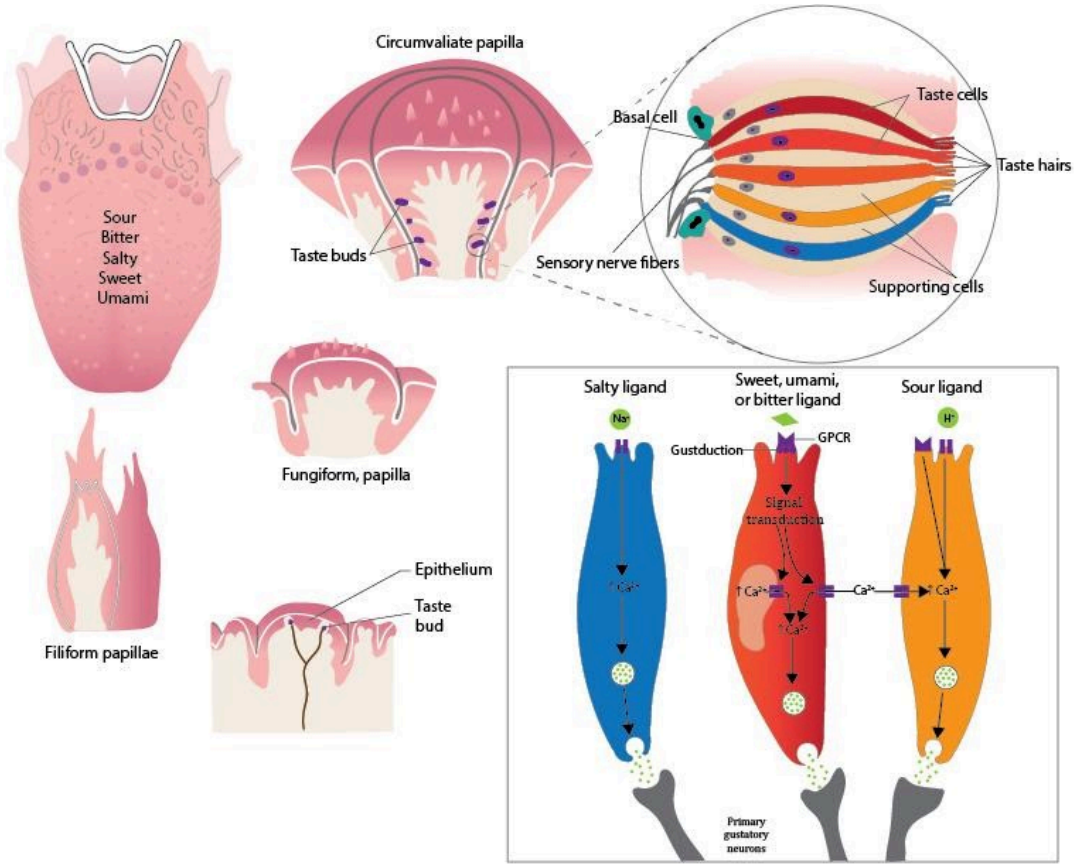


Illustration of a typical taste bud containing the taste cells.
© 2013 Encyclopædia Britannica, Inc. Downloaded from Image Quest Britannica; BYU-Idaho.

As with all sensory systems, taste receptor cells must somehow convert an external or environmental stimulus into an internal or cellular response. In the body, these cellular responses are often associated with changes in membrane potentials. When the receptor is a non-neural cell, as is the case with taste, the depolarizations are called **receptor potentials**. The receptor potential is not an action potential but a graded potential that can modulate the activity of ion channels and trigger an action potential in a neuron. Typically the neurons that are associated with taste cells have a resting membrane potential of around -70mV. In order for a response (an action potential) to be generated something has to cause the membrane potential of the neuron to reach threshold. In the case of taste, a neurotransmitter is released from the basal membrane of the taste receptor cell. Release of the neurotransmitter is triggered when a tastant interacts with the apical surface of the receptor cell. This interaction generates a depolarizing receptor potential that causes voltage-gated calcium channels on the basal surface of the cell to open. Calcium diffuses into the cell and acts as a signal to trigger exocytosis, releasing the neurotransmitter onto the neural cells. The neurotransmitter then elicits an action potential in the neuron. (Note: If you don't remember how a chemical synapse works, review module 6.) Even though different taste receptor cells respond to different tastants, the release of a neurotransmitter from the

receptor cells is always caused by an increase in intracellular Ca^{++}. Thus, to understand how a chemical tastant (external stimulus) results in a neuronal signal (internal response) we need to understand how the tastant can cause a depolarization of the receptor cell and thus an increase in intracellular Ca^{++}.

For a substance to be tasted, it must first dissolve in the saliva of the mouth. Once dissolved, the substance can interact with the apical membrane of a specific receptor cell. Receptor cells that respond to salt or sour substance contain membrane protein channels that allow the salty ion (Na^+) or the sour ion (H^+) to pass directly through the channel.



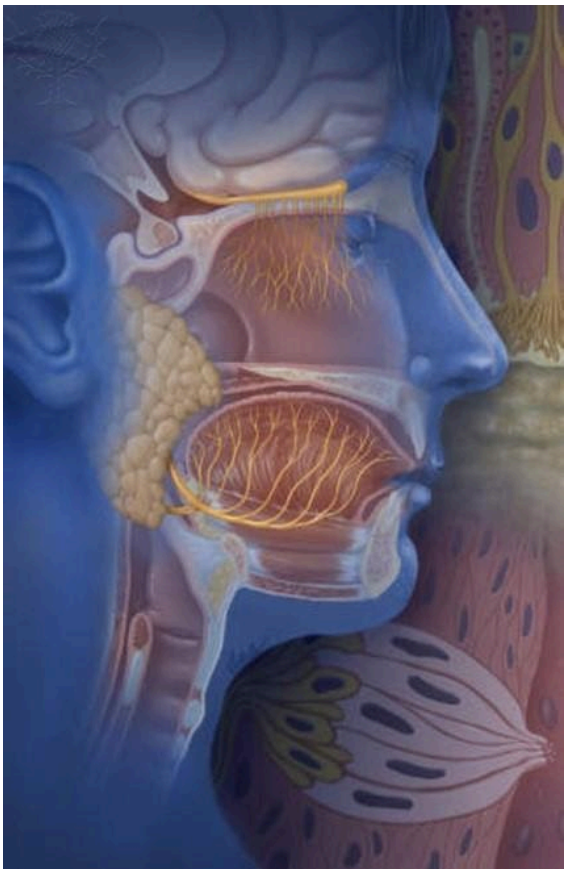
Signal transduction in a cell in response to a given taste stimulus. *Image by Jared Cardinet BYU-Idaho Fall 2015*

This increased movement of positively charged ions across the membrane causes depolarization which in turn causes voltage sensitive Ca^{++} channels to open. Sensations of bitter, sweet and umami all use G-protein coupled receptors, thus, these substances don't cross the membrane, but the effect is still the same (Ca^{++} increases in the cell which facilitates the release of neurotransmitter). Molecules that we perceive as bitter, sweet or “savory” (umami) bind as a ligand to G-protein coupled receptors (GPCR). This triggers a second messenger system that results in the opening of Ca^{++} channels on the surface of the taste receptor cell or on the smooth endoplasmic reticulum inside the cell (or both). Ca^{++} enters the cytosol and helps trigger the exocytosis of vessicles full of neurotransmitter.

At least three different properties of the taste sensation pathway allow the brain to distinguish between different taste sensations: (1) the proportion of action potentials received from the different receptor types, (2) additional input from the smell, and (3) additional input from other types of sensory receptors in the mouth. More specifically, foods that contain a higher proportion of salty to sweet chemicals will taste more salty. The sense of smell is very important to our perception of taste. Think of how food tastes when you have a cold and your nasal passageways are congested. Without the input of smell, many foods taste bland, for example, an onion tastes more like an apple. Other input from receptors that detect temperature, texture, and pain enhance the perception of the food. An example of this enhancement is the chemical **capsaicin**. This chemical is found in varying quantities in different kinds of peppers, endowing them with spiciness. However, our perception of spiciness is not the result of activation of a specific taste

receptor but is due to the activation of pain receptors. Still, this enhanced taste sensation from capsaicin is a must for many people.

The perception of taste is produced by neural processing within the primary sensory cortex of the brain as it receives action potentials from different nerve bundles. The three major nerves that contribute to gustation are the **facial nerve**, which receives input from taste receptors located on the anterior two-thirds of the tongue, the **glossopharyngeal nerve**, associated with the posterior one-third of the tongue, and the **vagus nerve**, associated with the surface of the epiglottis. All of the nerves converge on the solitary nucleus of the medulla oblongata and eventually arrive at the lateral regions of the primary somatosensory cortex.



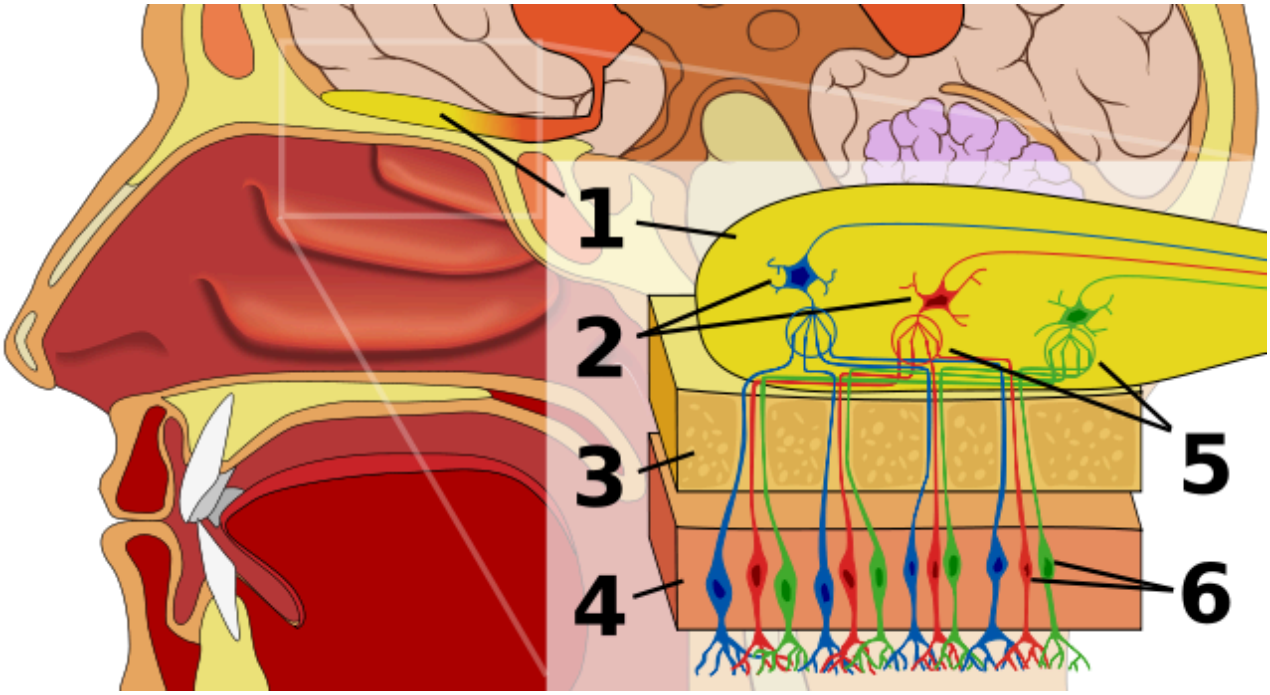
Olfactory Neurons, Facial Nerve and Glossopharyngeal Nerve.
© 2013 Encyclopædia Britannica, Inc. Downloaded from Image Quest Britannica; BYU-Idaho.

The image above shows the location of the facial and glossopharyngeal neurons innervating the tongue. The facial is the large neuron extending along the length of the jaw. The glossopharyngeal innervates the most posterior aspects of the tongue. This image also shows the olfactory neurons of the olfactory epithelium communicating with the olfactory bulb at the top. Olfaction is also very important for the experience of taste.

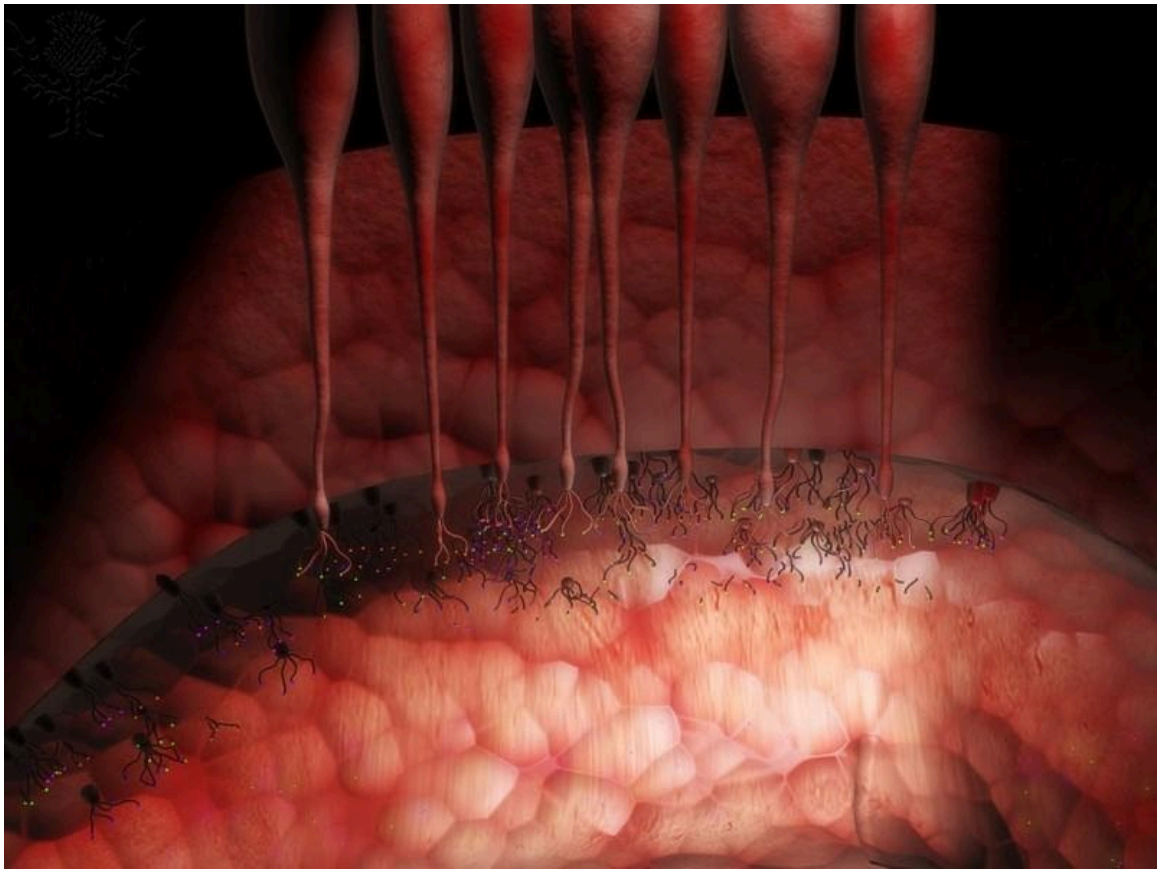
12.1.2

The Sense of Smell

Our other chemical sense is the sense of smell or olfaction. In contrast to taste cells, which are **epithelial derived**, **olfactory receptor cells are neurons**. (See the image above in 12.1). Similar to taste cells, olfactory receptor neurons have a relatively short life span, about 2 months, and must be replaced. The ability to grow in cycles (4 to 8 weeks) places olfactory neurons in a unique class within the nervous system as the only neurons that routinely die and are replaced. Olfactory receptor neurons are located within the olfactory epithelium, a 10 cm2 area located high in the roof of the nasal cavity. Animals with a more highly developed sense of smell have many more olfactory receptors than humans. For example, some dogs have an olfactory epithelium with 100 times as many olfactory receptor neurons covering an area of 170 cm2. Some breeds of dogs can detect the scents of people hours after the person has left the area. The title for all time best smeller goes to the bear. Black bears have been known to travel 18 miles in a straight line to a food source. African Elephants also have an excellent sense of smell and can detect water up to 12 miles away.



Olfactory Anatomy. *Title: File: Olfactory System.svg; Author: Chabacano;*
Site: https://books.byui.edu/-tYJc;
License: This file is licensed under the Creative Commons Attribution-Share Alike 2.5 Generic license.
The image above Shows (1) Olfactory bulb; (2) Mitral Cells; (3) Bone; (4) Olfactory Epithelium; (5) Glomerulus; (6) Olfactory neurons



Olfactory Neurons: Smell receptors. *Encyclopædia Britannica ImageQuest. Web. 19 Jun 2015. BYU-Idaho.*
Illustration of the location of the olfactory neurons (neurons at the top of the nasal cavity). These are bipolar neurons whose axons extend through the cribriform plate to the olfactory bulb. Note the cilia (olfactory hairs) extending into the mucous lining the olfactory epithelium.

The olfactory epithelia contain millions of olfactory sensory neurons that can detect more than 400,000 different substances, by some estimates, and more than 1.7 trillion, yes 1.7 trillion different smells (Bushdid et al., 2014; Science). As we breathe, chemical odorants are mixed with the air as they pass through the many folds in the nasal passages. In order for us to smell, the odorant must first be dissolved in a thin water-based mucous layer that lines the olfactory epithelium. Dissolved odorants then bind to receptors located within membranes of the receptor cilia of the olfactory neuron (see image above). Olfactory receptors are G protein coupled receptors that activate the enzyme adenylyl cyclase. The activated adenylyl cyclase, in turn, produces cAMP which opens a cation channel, allowing Na+ and Ca++ to enter the cell. The influx of the positively charged ions depolarizes the membrane triggering an action potential.

The olfactory sensory neurons synapse with neurons in the olfactory bulb that extend to the olfactory cortex. Neurons from the olfactory tract pass either to the olfactory epithelium in the temporal lobe where the odor is perceived or to areas in the frontal lobe where emotions associated with the odor are generated. These areas are associated with the limbic system and can generate emotional and visceral responses to the smell. Recall that the sense of smell is the only major sense that does not pass through the thalamus. It is considered to be a very primitive sense and is important in emotion and memory. You may have observed an increase in emotional responses to smells during pregnancy. It is thought that the hormones of pregnancy enhance the cortical regions of smell so that the woman becomes more sensitive to certain odors.

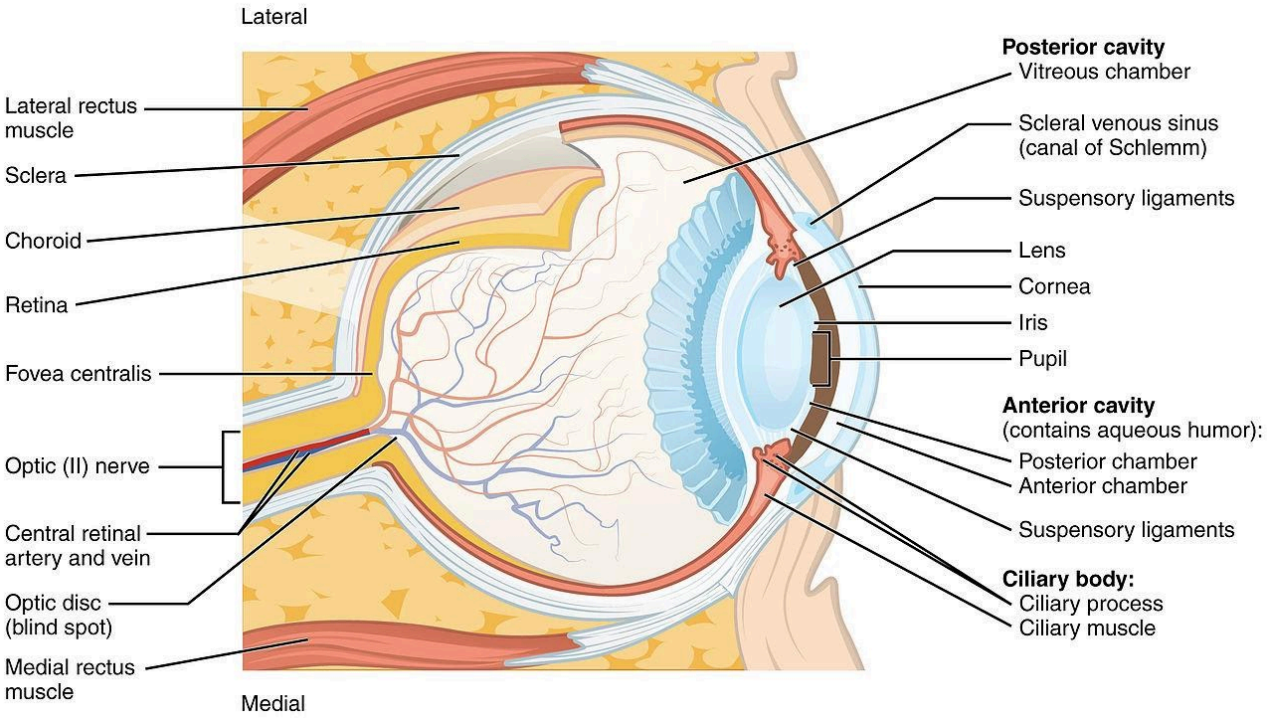
12.2

VISION: STRUCTURE OF THE EYE

12.2.1

Anatomy of the Eye

The eye is a hollow, fluid filled organ that is surrounded by three layers of tissue (see image below). The outermost layer, the **avascular tunic**, is composed of connective tissue. As the name implies there are no blood vessels penetrating this layer. It can be divided into two parts, the **sclera**, the white part of the eye comprising the posterior 5/6 of the eyeball, and the **cornea**, the clear window on the anterior surface of the eye. The sclera helps protect the eye and also provides a site of attachment for the six muscles responsible for the movement of the eye. The cornea is transparent and functions as the major refractor of the light as it enters the eye. Its transparency is due to the nature of the collagen and proteoglycan fibers that form it. Following are a couple of pictures that help orient us to the anatomy of the eye. It may help to print these and have them in hand as you read the following sections.



Structure of the Eye. *Title: File:1413 Structure of the Eye.jpg; Author: OpenStax College; Site: https://books.byui.edu/-Yfw;*

License: This file is licensed under the Creative Commons Attribution 3.0 Unported license.

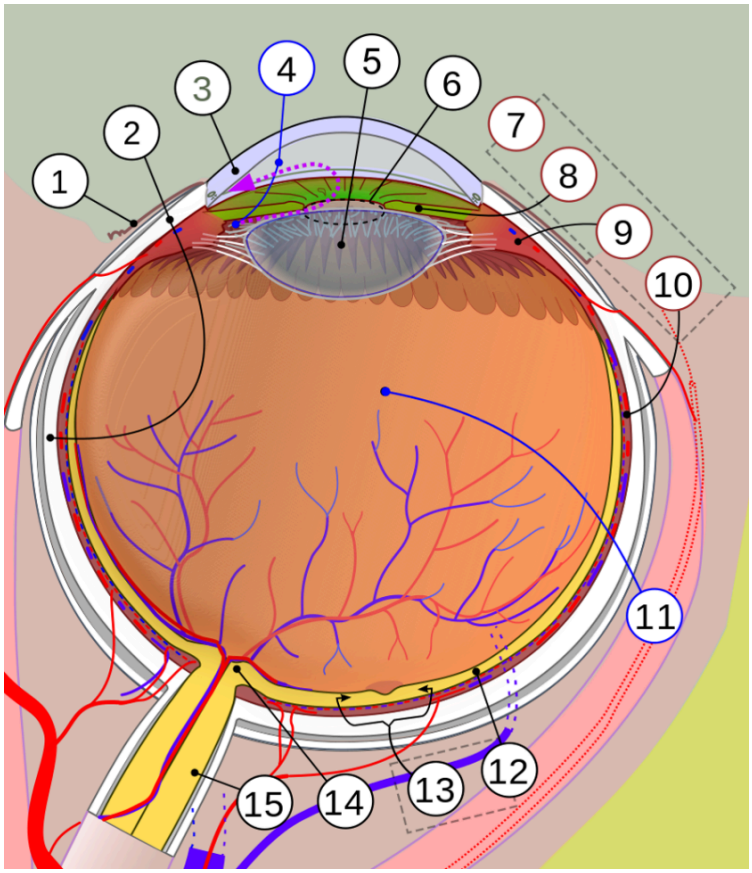


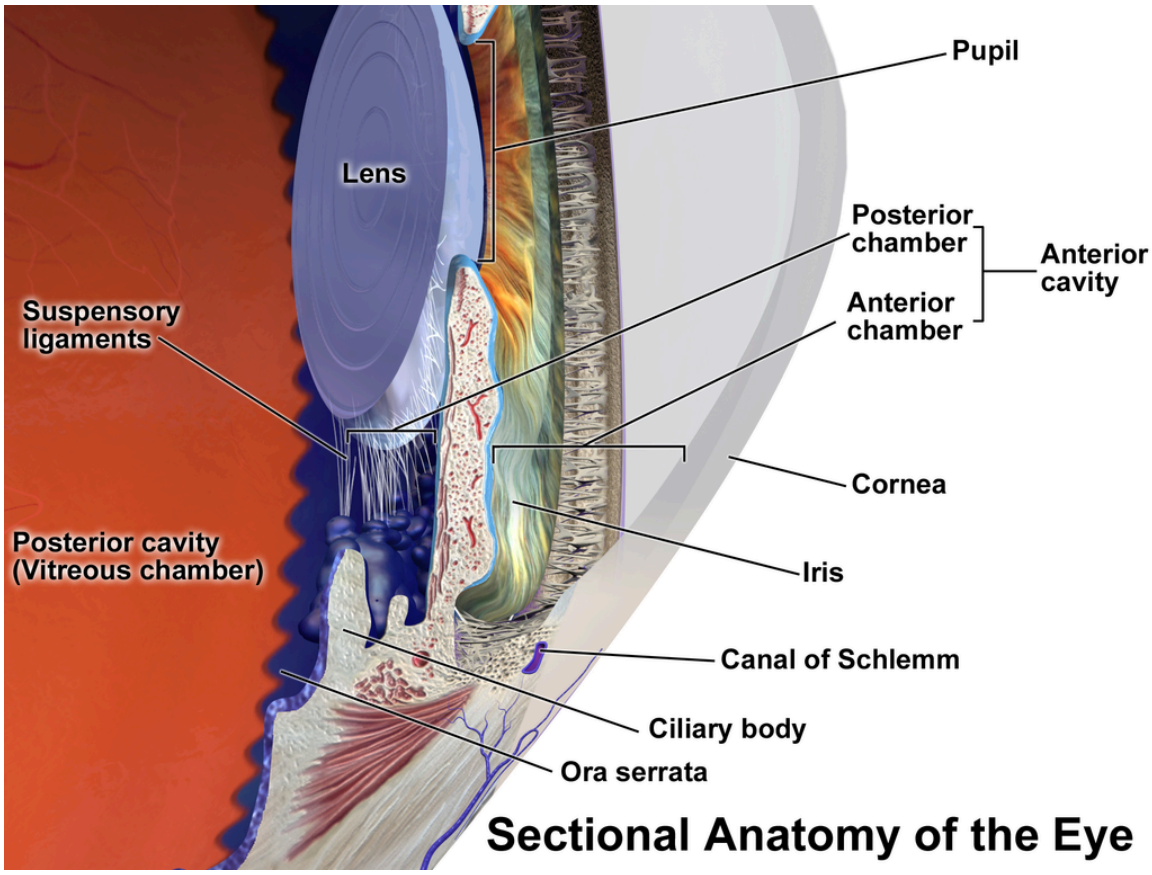
Diagram of Human Eye. *Title: File: Simple diagram of human eye multilingual.svg; Author: Jmarchn; Site: https://books.byui.edu/-WHDm;*

License: This file is licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license.

1. Conjunctiva
2. Sclera
3. Cornea
4. Aqueous humour (in anterior and posterior chambers. See purple dotted line)
5. Lens
6. Pupil
7. Uvea with
8. Iris
9. Ciliary body and
10. Choroid
11. Vitreous humor
12. Retina with
13. Macula or macula lutea
14. Optic disc → blind spot
15. Optic nerve

The middle layer of the eye is the **vascular tunic**. Most of the blood vessels of the eye can be found in this layer. The picture above shows blood vessels of the retina. The blood vessels of the vascular tunic are not shown. If they were shown, you would see them associated with the choroid in the image. The posterior portion of this layer is the choroid. Anteriorly the choroid is continuous with the **ciliary body**. The ciliary body is composed of a ring of smooth muscle, the **ciliary muscle**, and the **ciliary processes**. The ciliary muscle is a sphincter-like muscle that is attached to the lens

capsule via the suspensory ligaments. It is responsible for adjusting the thickness of the lens. The ciliary processes are secretory structures that produce the **aqueous humor** that fills the compartment in front of the lens.



Anatomy of the Eye. *Title: Blausen 0390 EyeAnatomy Sectional.png; Author: BruceBlaus; Site: https://books.byui.edu/-vDgy;*

License: This file is licensed under the Creative Commons Attribution 3.0 Unported license.

The most anterior part of the vascular tunic is the iris. The iris is composed primarily of smooth muscle containing varying amounts of the pigment melanin. The amount of melanin determines eye color, large amounts produce brown eyes, while smaller amounts result in blue or green eyes. The iris is actually two layers of muscle with a circular hole in the center, the **pupil**. The **sphincter pupillae** is a circular layer that causes the pupil to constrict (**miosis**) when it contracts and the **dilator pupillae** is a radial layer that causes the pupil to dilate (**mydriasis**) when it contracts (see image below). These layers are innervated by the autonomic nervous system, the dilator is under sympathetic control and the sphincter is under parasympathetic control.

Sphincter Pupillae: Dilation to Constriction.

Photograph of the eye: From left to right we have normal, mydriasis, and miosis.

The innermost layer is the **neural tunic** or **retina**. There are actually two distinct layers of the retina. The **pigment epithelium** is a layer of simple cuboidal epithelium that sits on the choroid. This layer has large amounts of melanin giving it a dark black color. One important function of the pigmented retinal is to absorb light that doesn't strike the photoreceptors and prevent it from being reflected inside the eye. The **neural layer** is the innermost layer of the wall of the eye and contains the photoreceptors that are stimulated by the entering light. Two distinct anatomical structures on the retina are the **optic disk** and the **fovea centralis**. The optic disk, also called the blind spot, is the point where the optic nerve and blood vessels enter the eye. There are no photoreceptors in this area and hence light striking the optic disk cannot be detected. The fovea centralis (fovea = pit) is a small indention located in the center of a special area of the retina called the **macula lutea** (macula = body, lutea = yellow). The macula is roughly 5 mm in diameter, about the diameter of a pencil eraser, and the fovea is about the size of the head of a pin. When you look at an object the light coming directly from that object focuses on the fovea, it is the portion of the retina with the greatest visual acuity (clarity).

The **lens** is not technically part of any of these three layers but it is obviously extremely important in focusing light. It is a biconvex structure composed of transparent cells (epithelial cells). These cells have lost their nuclei and other organelles and are filled with transparent proteins called **crystallines**. It is surrounded by the very elastic lens capsule which, in turn, is attached to the ciliary muscles by the suspensory ligaments. When there is no tension on the suspensory ligaments (ciliary muscles are contracted) the lens assumes its natural shape, this is when it is at its thickest. When the ciliary muscles relax the tension on the suspensory ligaments increases and the lens flattens. Remember the ciliary muscle is a sphincter muscle so when it contracts its diameter decreases, reducing tension on the ligaments attached to the lens capsule.

The lens divides the eye into two fluid filled compartments. The **anterior cavity** is the space between the lens and the cornea. As was mentioned above, this cavity is filled with the **aqueous humor** produced by the ciliary processes. Aqueous humor is a watery fluid produced continually and circulates through the cavity before being reabsorbed into the blood. It is important in maintaining proper intraocular pressure as well as circulating nutrients and removing wastes to the cells of the lens and cornea. If the normal circulation is blocked it can result in an inappropriate increase in pressure, a condition known as **glaucoma**. If not treated, glaucoma can result in vision loss and blindness. You may see some anatomy texts divide the anterior cavity into two "chambers." The anterior chamber of the anterior cavity would be between the cornea and the iris. The posterior chamber of the anterior cavity would be a very small space between the iris and the lens. The **posterior cavity** is the space behind the lens. This compartment is filled with **vitreous humor**. Vitreous humor is more of a gel, similar to egg white. It also is important in maintaining intraocular pressure, but unlike aqueous humor, turns over very slowly.

12.2.2

Focusing Light on the Retina

In order to clearly see any object, the first thing that has to happen is that the light reflected off of the object must be focused on the retina. Think of the light being reflected off of a face that you are looking at. Light coming from the subject's nose is spreading out (diverging) as it hits your eye. In order to see the nose clearly, all of the light reflecting from it must be focused on a single spot on the retina. To accomplish this, the light needs to be bent or refracted. You know that light can be focused using something like a magnifying glass. The physics behind this phenomenon has to do with the fact that as light passes through objects of different densities its speed changes. If it strikes the object at a 90-degree angle, even though its speed changes it maintains a straight path. However, if it strikes the object at any other angle (not 90 degrees) it is refracted (see image below). This is why a convex magnifying lens can focus light. We have two convex structures in the eye to bend the light, the cornea and the lens. Also, their densities are different than the air and they are different than the aqueous and vitreous humors. We, therefore, have the ideal conditions for bending the light. However, depending on how close the object is to our eyes the light will be diverging at different angles. The closer it is, the more the light diverges and the more it must be refracted to focus properly. This requires that we be able to adjust the curvature of our refracting surfaces in order to focus properly. The cornea is an excellent refractor, in fact it is responsible for most of the required bending of the light. Its limitation, however, is that it cannot be adjusted. The lens, on the other hand, is adjustable due to its elasticity and the actions of the ciliary muscle. We define a surface's refracting ability based on its focal point. The focal point is the precise point at which the light rays all converge. As the lens becomes thicker its focal point shortens and it is able to bend light more sharply. On the other hand, as the lens flattens its focal point becomes longer and it bends the light to a lesser degree. Let's see if we can put this all together and explain how we focus objects at different distances from the eye.

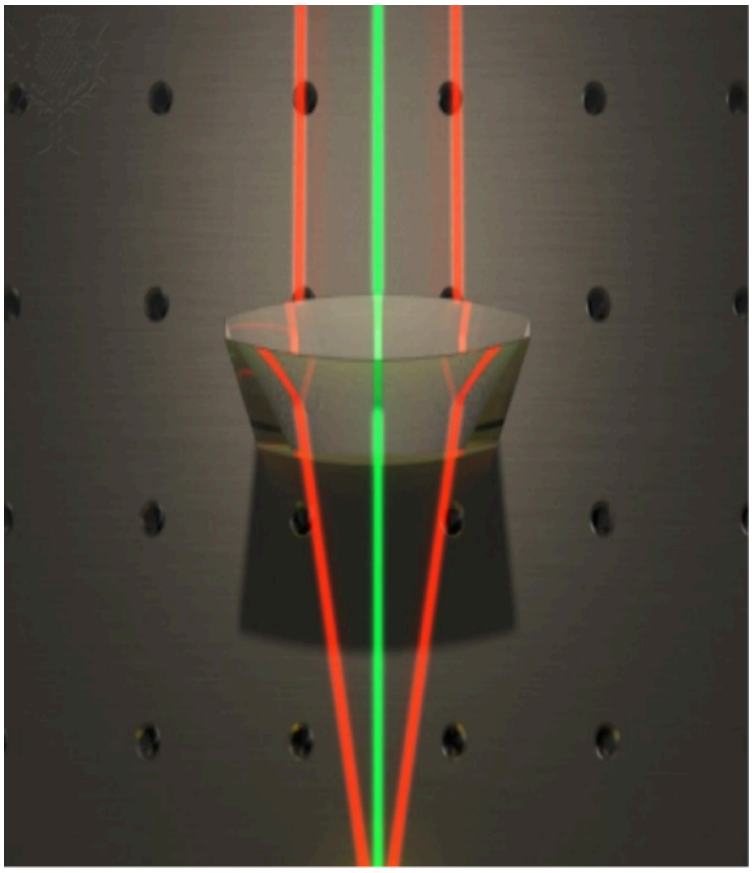


Illustration of light refraction through a prism which causes the light rays to bend and converge at a single focal point. *© 2013 Encyclopaedia Britannica, Inc. Downloaded from Image Quest Britannica, BYU-Idaho*

We will start with objects that are 20 feet or more from the eye. At this distance the normal eye is designed to focus the object properly without any thickening of the lens. The ciliary muscles would be relaxed and the lens would be at its thinnest. The closest distance at which the lens does not have to thicken for proper focusing is called the **far point of vision**. For the normal, average eye, the far point of vision is 20 feet. For anything less than this distance from the eye, three things must happen to properly focus the object on the retina. The first we have already discussed; the lens must thicken. This phenomenon is known as **accommodation**. As objects continue to move closer the lens will thicken more and more until it is at its maximum thickness. If the object is brought even closer it will begin to blur. The closest point that at which we can keep the object in focus is called the **near point of vision**. The near point of vision changes as we get older. It is only 2-3 inches in infants but may be as far as 5 feet when we get into our late 40's. The change is due to the fact that the lens becomes less elastic as we age and cannot thicken as much, a condition known as **presbyopia**.

Focus on the back wall of your room and while maintaining your focus bring your finger in front of your nose a few inches from your face. What do you see? The reason you see two images is because the light is focusing on different parts of the retinas in your two eyes. Now focus on your finger so that you only see one. In order to do that your eyes had to turn in or converge. The closer the object is to the eyes, the more they have to **converge**. This is the second thing that must happen in order to properly focus on objects less than 20 feet away.

The third thing that happens is that the pupils constrict. The purpose for this constriction is to increase the **depth of focus**. The depth of focus is how much of the visual field we can keep in sharp focus. Again, place your finger between your face and the computer screen. If you focus on your finger the print on the computer screen will blur and if you focus on the print your finger will blur, we cannot keep both in focus at the same time. Constriction of the pupil increases the depth of focus and helps us keep close objects totally in focus. This is also why we sometimes require additional light to see clearly when doing really close work. Since the pupil is constricted, less light can enter the eye requiring more light to see well.

Focusing Errors

Even though our eyes are designed to focus automatically, like a self-focusing camera, problems do arise. Think of how many people you know who have to wear glasses or contact lenses in order to see clearly. The most common focusing error is **near sightedness, or myopia**. People who are near sighted can see fine up close but have a difficult time focusing things in the distance. This is usually due to an eyeball that has grown too long. Recall that the eye is designed to properly focus objects greater than 20 feet away without any accommodation of the lens. If the eyeball is too long the image focuses in front of the retina (the focal point is too short) causing objects to appear blurred in our vision. To correct this condition, lenses that spread the light are used to lengthen the focal point and achieve proper focusing. Concave lenses spread light. If you are near sighted and wear glasses, note that the lens of your glasses is thinner in the center than on the edges, creating the concave lens to spread the light. A simple eye test is used to determine if you are nearsighted. This test involves a chart with lines of letters that get progressively smaller as they go down the chart. The subject stands 20 feet from the chart and reads the smallest print that he can. He is then assigned a number based on which line he can read. Normal vision is 20/20 vision. These numbers represent distances. The first number is where the subject is standing, i.e. 20 feet from the chart. The second number is where someone with normal vision would stand to read the same line as the subject. For example, if you have 20/20 vision you can see at 20 feet what a “normal” subject would see at 20 feet. If your vision is 20/80 it means that what you see at 20 feet, a “normal” subject would be able to see at 80 feet.

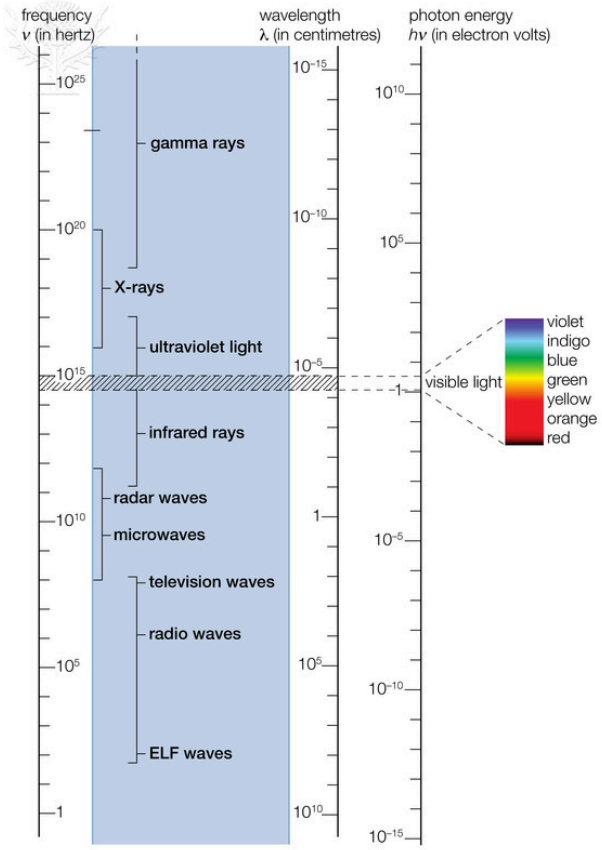
Another vision problem, **far sightedness, or hyperopia**, is less common and is essentially the opposite of myopia. The subject can see distant objects well but close objects are blurred. The usual cause of this disorder is an eyeball that is too short so the lens cannot thicken enough. Consequently, the object focuses behind the retina. To correct for farsightedness, a convex lens is used to shorten the focal length. Children with hyperopia will sometimes grow out of the problem when their eyeball lengthens as they age.

Two other common conditions are **astigmatism** and **presbyopia**. Astigmatism is due to an irregularity in the lens or the cornea such that one or both is not symmetrical. This results in part of the object focusing normally while part of it is out of focus. To correct this condition lenses are prepared that are also asymmetrical to counteract the irregularities in the lens or cornea. Presbyopia was alluded to earlier. This is the condition that develops as we age and the lens becomes less elastic. Often, older individuals who have always had normal vision will find that they need reading glasses to help focus the light. The corrective lenses for presbyopia are the same as for far sightedness. They are convex to shorten the focal point of the light.

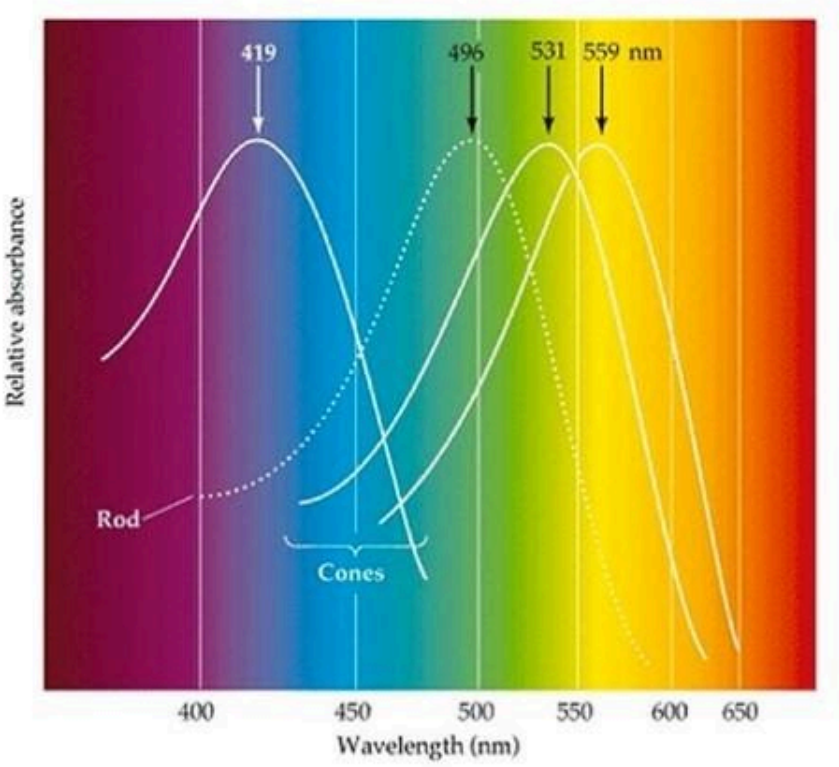
12.3

CONVERTING LIGHT TO ACTION POTENTIALS

The function of the eye is to convert light waves to action potentials. In order to understand how this happens, we need to know a little about the nature of light. Visible light is a very small portion of the spectrum of **electromagnetic radiation**. The entire spectrum of electromagnetic radiation is shown in the image below.



Spectrum of Electromagnetic Radiation. The visible spectrum is shown as colors.



Light Spectrum Wavelength and Peak Absorption from Cone and Rod Cells.
Title: Spectrum.jpeg; Author: http://webvision.med.utah.edu/;
Site: https://books.byui.edu/-Yjww;License: Copyright © 2015 Webvision: Attribution, Noncommercial, No Derivative Works Creative Commons license.
The image above shows the peak absorption of each of the cone cells as well as the rod cells.

The nature of electromagnetic radiation, and hence visible light, cannot be described using a single model. Some of light's properties can be explained by describing it as a wave. For example, the color of light that we perceive is based on the **wavelength** of the light waves. However, other properties suggest that light exists as discrete packets of energy called **photons**. The image above shows the relationship between wavelength and the energy in a photon of light. The shorter the wavelength, the greater the energy. Hence, gamma waves have very short wavelengths and contain large amounts of energy while radio waves have very long wavelengths but relatively small amounts of energy. The portion of the spectrum of electromagnetic radiation that we can perceive is referred to as the **visible spectrum** and includes light with wavelengths between 380 (violet) and 700 nm (red).

When light strikes an object, one of three things will happen. If the object is transparent the light is **transmitted**, meaning it will pass through the object. However, if the object is not transparent the light will either be **absorbed** or it will be **reflected**. The color that we perceive as we look at an object is due to the light that is being reflected off of it. Hence, if we see yellow, the yellow wavelength light is being reflected and the other wavelengths are being absorbed. Objects that appear black to our eyes absorb all of the light that is striking them while objects that appear white reflect all of the light that is striking them.

The Retina

Phototransduction

12.3.1

The Retina

The structure of the eye responsible for converting light waves into action potentials is the retina. The neural layer of the retina is composed of three main types of cells: the **photoreceptors**, the **bipolar neurons** and the **ganglion cells**. The photoreceptors, as the name implies, have the responsibility of capturing the light and converting it to an electrical signal. There are two types of photoreceptors in the retina, the **rods** and the **cones**. The rods see only in black and white and are mainly responsible for our night vision. The cones, on the other hand, can see in color and are responsible for color vision as well as sharp vision. Each eye contains about 120,000,000 rods and 6,000,000 cones. Although they detect light of different wavelengths, structurally, rods and cones are similar. They are composed of an **outer segment** that touches the pigment epithelium and is composed of numerous flattened discs stacked on each other (think of a stack of dinner plates). The only difference is that in the rods, all of the discs are the same size while in the cones they gradually decrease in diameter as they move to the end of the cell. This results in the shape for which the cones were named. The outer segment connects to the **inner segment** which houses the nucleus and other organelles of the cell. The inner segment, in turn connects to the **synaptic terminal** which forms the connection between the photoreceptor and the bipolar neuron (see the images below).

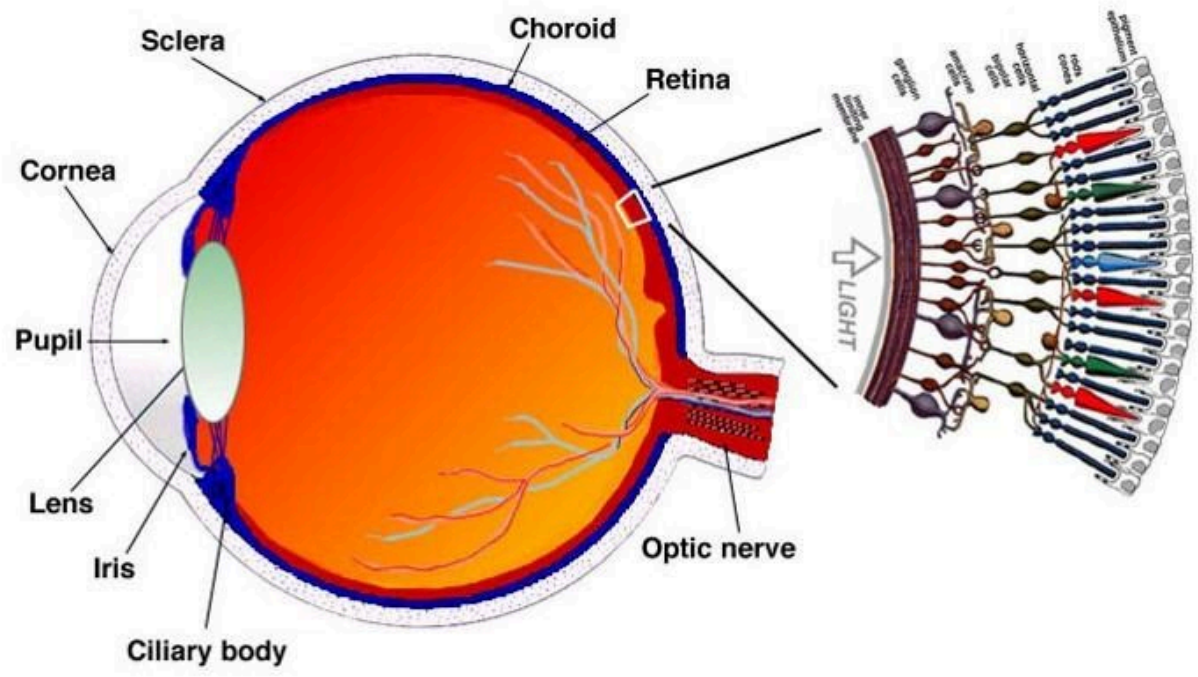


Fig. 1.1. A drawing of a section through the human eye with a schematic enlargement of the retina.

Cross section of Human Eye and Retina Enlargement. *Title: Sagschem.jpeg; Author: http://webvision.med.utah.edu/; Site: https://books.byui.edu/-MajL; License: Copyright © 2015 Webvision: Attribution, Noncommercial, No Derivative Works Creative Commons license.*

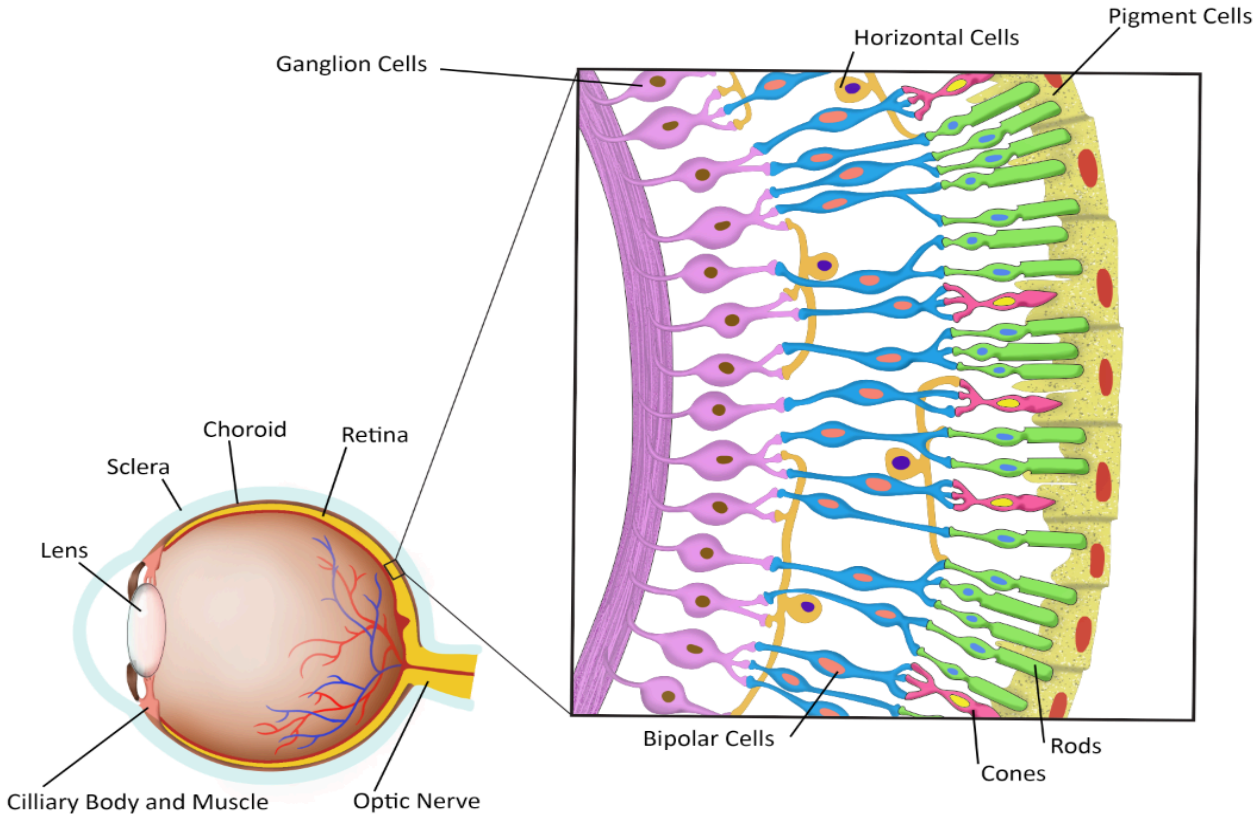


Image by Becky T. BYU-I W20.

Illustration of the photoreceptors of the eye, rods (black) and cones (green) and associated neurons within the neural retina. The brown cells are the bipolar neurons and the large orange structures are the cell bodies of the ganglion cells. The red colored layer at the top represents the choroid and the top purple layer the sclera.

The bipolar neurons, so named because they have one dendrite and one axon, are the connections between the photoreceptors and the ganglion cells. The axons of the ganglion cells form the optic nerve which exits the eye via the optic disc. Two other cell types are shown in the image above. The **horizontal cells** can be seen in the layer where the photoreceptors synapse with the bipolar neurons and the **Amacrine cells** can be seen in the layer where the bipolar cells synapse with the ganglion cells. These two cells are involved in modulating the visual signals.

The distribution of the photoreceptors in the retina is not uniform. In the fovea centralis we find only cones. Moving away out from the fovea we start to see rods intermixed with the cones. The further out from the fovea we move the greater the number of rods and the fewer the number of cones.

In the image above note that the photoreceptors are located at the back of the retina. Light entering the eye must pass through the ganglion cells and the bipolar neurons before it gets to the photoreceptors. This doesn't seem to be the best arrangement. However, at the fovea, the ganglion cells and the bipolar neurons radiate away from the cones in the fovea. Think of the crown of your head. All of the hairs radiate out from this point exposing the scalp. Because of this arrangement light striking the fovea has direct access to the photoreceptors, enhancing vision in this region of the retina.

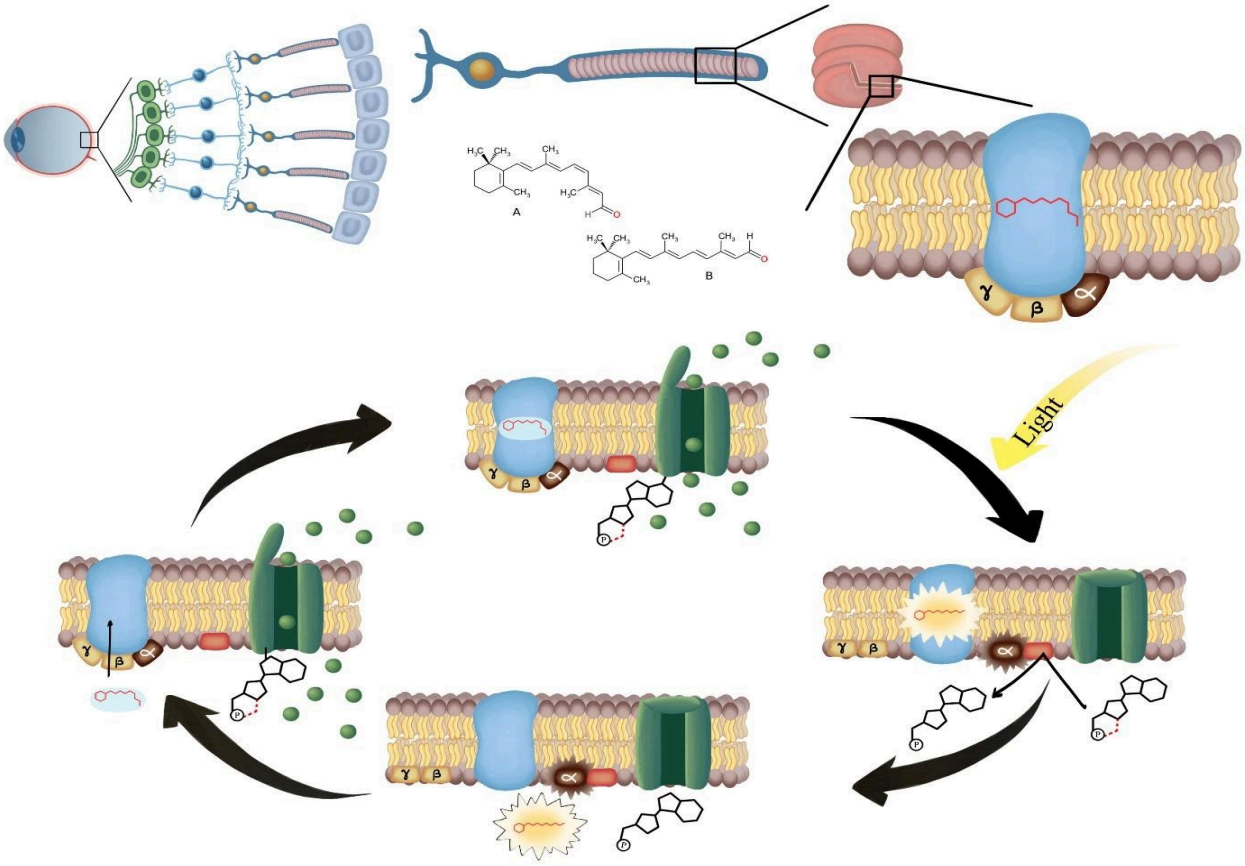
Now let's examine the unique characteristics of the different photoreceptors, starting with the rods. The rods are very

sensitive to light and will respond to a single photon of light. In addition, they are part of convergent circuits in which several rods will converge on a single bipolar neuron and several bipolar neurons will converge on one ganglion cell. This allows for the summation of signals from several rods resulting in an action potential being sent to the brain. These properties make the rods ideal for seeing in very dim light, therefore rods are responsible for our night vision. During the day when there is plenty of light, the rods are essentially inactivated due to a process called bleaching (more on this later). You are aware of how hard it is to see in a darkened theater when you first enter from bright light. After you have been there for a while and the rods become active we can see quite well in the room. Indeed, after about 40 minutes in the dark room our eyes are about 25,000 times more sensitive than they were when we first entered the room. There are two downsides to the use of rods, however. First, they do not see in color, rods see only in black and white. Second, due to the convergence their visual fields are quite large. Light striking any of the rods that converge on one ganglion cell will produce the same “pixel.” Therefore, vision with rods is very sensitive but not very acute (sharp). Cones, on the other hand, have essentially the opposite characteristics. First, there is very little convergence in their circuitry. Light striking two cones located next to each other would produce two different pixels in the brain. This allows for very sharp (acute) vision for images striking the fovea since there are only cones in the fovea. Second, the cones are much less sensitive than rods. At night, the intensity of light usually is not sufficient to stimulate the cones. Third, the cones are responsible for our color vision. We have three types of cones that respond to light in the red, green or blue wavelengths. By mixing the input from these three cones, humans can perceive about 1,000,000 different hues of color. You may know someone who is “color blind.” Color blindness is most often due to a genetic condition where the subject does not produce one or more of the cones. The most common condition is red-green colorblindness, where the person lacks either the red or green cone. Individuals with this condition can see colors, but they have a difficult time distinguishing between shades of green and red. The genes for the green and red cones are found on the X chromosome, therefore males have a much higher incidence of color blindness since males only have one X chromosome. Women have two X chromosomes so even if they inherit a defective gene there is a good change the gene on the other X chromosome will be normal. For a women to be red-green colorblind, both her father and her mother would have to have the condition. Another type of color blindness called Blue-yellow color blindness involves genes for the blue cones, but these genes are not on the X chromosome so it occurs at the same rate in both males and females.

12.3.2

Phototransduction

Now for the underlying question, how do the proteins that absorb photons of light produce the action potentials that travel to the brain to produce what we perceive as vision? This process is called **phototransduction** (see figure below). Since the process is essentially the same in both the rods and the cones we will look at the rods and then explain the subtle differences that occur in the cones. It all starts with the visual pigments that are embedded in the membranes of the disks found in the outer segment of the rods. This visual pigment is called **rhodopsin** and is composed of protein called **opsin** and a derivative of vitamin A called **retinal**. In the unexcited state, retinal has a bend in its hydrocarbon chain (11-cis retinal) and fits nicely in a binding site on the opsin. When light of the proper wavelength is absorbed by the visual pigment the energy of the light causes retinal to change shape and the hydrocarbon chain loses its bend (all-trans retinal) and no longer fits in the binding site. It should be noted that even though rods provide only non-color vision, light of the green wavelength is the most efficient in activating rhodopsin. When the retina detaches from the opsin it becomes inactive. This process is known as **bleaching**. Opsin is actually a G-Protein coupled receptor (GPCR) that is activated by light, hence it is a photoreceptor. Once activated the GPCR activates the G-protein, separating the alpha subunit from the beta/gamma subunit (see module 5 for a review of GPCRs). In the photoreceptors of the eye the G-protein is called **transducin**. The alpha subunit then brings about a change in the cell. More on this later.



Phototransduction. *Created by BYU-Idaho student Hannah Crowder, 2013*
Phototransduction. The top image represents the photoreceptor in the dark. The green channel is the cGMP gated cation channel which is open and allowing cations (Na^+ and Ca^{++}) to depolarize the cell. When light strikes and changes the retinal from 11-cis to all-trans retinal it activates the G-protein transducin which results in the breakdown of cGMP and the closing of the cation channel. The cell will then hyperpolarize. Finally, All-trans retinal is converted back to 11-cis retinal and it re-attaches to opsin allowing cGMP to open the cation channel and once again depolarize the cell.

Photoreceptors are different than any receptors we have discussed to date in that they release neurotransmitter when they are **not** being stimulated. Here is how this works. There are three important ion channels in the membranes of the photoreceptor cells, **K^+ leak channels, voltage-gated Ca^{2+} channels** and **cyclic GMP (cGMP) gated cation channels** (Na^+ and some Ca^{2+} move through this channel). When the photoreceptor is not being stimulated (in the dark), cGMP is bound to the cation channel and Na^+ and Ca^{2+} diffuse into the cell maintaining it in a depolarized state. This depolarization causes the voltage gated Ca^{2+} channels to open, allowing more Ca^{2+} to diffuse into the cell. This Ca^{2+} triggers the release of the neurotransmitter glutamate by the process of exocytosis. The binding of glutamate to receptors on the bipolar neurons may be excitatory or inhibitory; it depends on what receptors are expressed on the bipolar neuron. In this module, we will focus on just the bipolar neurons that express receptors that cause **inhibition** when glutamate is attached.

When light is absorbed by rhodopsin and the G-protein (called transducin) is activated, the alpha subunit of the G-protein activates the enzyme **phosphodiesterase**. This enzyme breaks down cGMP to GMP. Once the cGMP is removed the cGMP-gated cation channels close and the membrane hyperpolarizes. This results in the closing of the voltage-gated Ca^{2+} channels and glutamate release ceases. Removal of the inhibitory signal to the bipolar neurons allows them to fire and an action potential is sent to the brain. Eventually, the G-protein is inactivated and phosphodiesterase is turned off. However, the rhodopsin cannot respond to light again until the retinal is returned to its bent, 11-cis, state. To do this, it diffuses into the pigment epithelium where enzymes act to restore the 11-cis state. It can then diffuse back into the rod cell and bind to opsin. The rod cell is ready to be activated again. The original bleaching process is very fast, fractions of seconds, but restoring the rhodopsin to its intact state can take several minutes. During the day, when we are exposed to sufficient light, the rhodopsin remains in the bleached state and the rods are essentially unresponsive to light. The mechanism is similar in the cones. The main difference is in the proteins of the visual pigment. The visual pigments in cones are similar to rhodopsin but they respond to different wavelengths of light allowing us to perceive different colors. Another difference, as stated above, is that the cones are much less sensitive to light. This is why the cones do well in full daylight when everything is brightly illuminated. Finally, cones do not stay deactivated (bleached) as long as rods. Cones appear to be fairly resistant to large scale “bleaching” as they are able to recover 11-cis-retinal much more quickly so that at any given time there are at least some visual pigments ready for stimulation.

Recall that the axons of the ganglion cells form the optic nerves. These nerves enter the brain through the optic canals of the skull. As they move posteriorly they converge at a point just above the hypothalamus called the optic chiasm (the word chiasm comes from the Greek letter Chi or X, implying a point of crossing over). In the optic chiasm some of the axons cross to the opposite side of the brain while some stay on the same side. Let’s see if we can make sense of this. If we use the fovea as a reference point we can divide the retina into two halves, a lateral or temporal half and a medial or nasal half. Axons originating on the lateral retina enter the optic chiasm but do not cross over while those from the medial retina cross over to the opposite side of the brain. What are the implications of this? Suppose you are looking straight ahead and light from an image at your right enters your eye. It will be focused on the lateral retina of your left eye and the medial retina of your right eye. Since axons from the lateral retina do not cross over while axons from the medial retina do cross over, the image will be projected to only the left hemisphere of your occipital lobe. An object to your left would be projected to only your right hemisphere and the object directly in front of you would go to both hemispheres. The overall result of this interesting circuitry is that it tells us how far away the objects are, in other words, it is responsible for our depth perception. Try closing one eye and judging how far you are from an object. You can do it but it is more difficult and less accurate.

From the optic chiasm, the optic nerves project to the thalamus where they synapse with the neurons that connect to the primary optic cortex in the occipital lobe of the cerebrum where it is perceived as an image. It is interesting that what we perceive isn't always what our eyes see. For example, as you gaze around the room everything seems like it is in sharp focus. The reality is that our eye is only capable of producing sharp vision on a very small portion of our visual field. If you hold your thumb at arm's length in front of you, the area covered by your thumbnail is about all the eye can focus sharply. Why then does everything seem clear? It is because our brain makes us think it is clear. Try focusing on something and then pay attention to the things on either side. They will not be in sharp focus but you didn't notice that until you thought about it. In reality, much of what we see is a product of our brains and not necessarily what the eye is seeing. For proof of this statement watch or listen to the TED talk below about, but beware they may blow your mind.

https://books.byui.edu/-Rxv (Transcripts available with videos website)

12.4

THE INNER EAR: SENSE OF HEARING AND EQUILIBRIUM

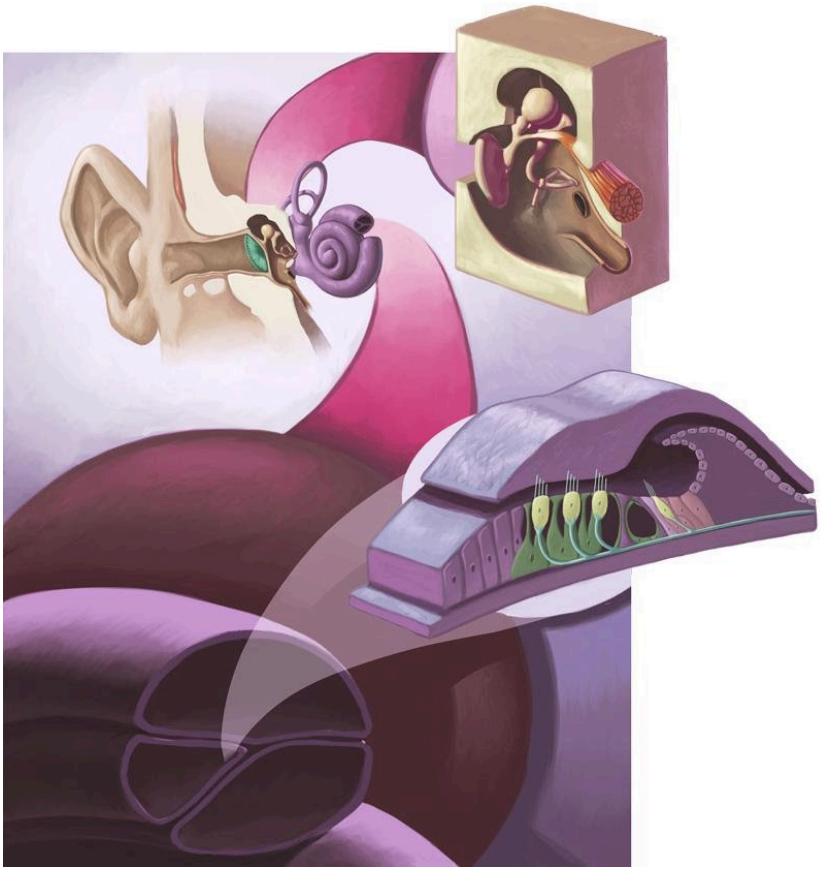
12.4.1

The Nature of Sound

The ear is a sensory organ designed to capture, transmit and translate sound waves into action potentials. Before we can understand how this is done we need to have a working knowledge of what sound waves are. Sound waves are pressure waves with alternating regions of compressed air and non-compressed (rarefied) atoms. Typically these waves occur in air, but they can also occur in other media such as fluids, and even solids. We will focus on the sound waves in the air since this is the usual means by which they enter our ears. We often depict sounds waves visually as an undulating line with the peak of the wave representing the compressed air and the valley the rarefied air. What we perceive as sound is a translation of the **frequency** and **amplitude** of the sound waves entering the ear. The frequency of sound is expressed as the number of peaks of the sound wave that pass a stationary point each second. Audible frequencies range from 20 **Hertz** (Hz) to 20,000 Hz (Hertz = cycles or peaks per second). Humans are most sensitive to sounds in the 1000-3000 Hz range. Our sense of hearing perceives different frequencies as different pitches. For example, low-frequency sound waves are perceived as low-pitched sounds and high-frequency sound waves are perceived as high-pitched sounds. The amplitude of the sound wave is the degree of compression and rarefaction. Visually we would depict sounds of higher amplitude as greater undulations in our line. Perceptually, amplitude determines the loudness of the sound. The loudness of a sound is measured in **decibels** (dB). A whisper measures about 30 dB while normal conversation measures about 60 dB. The decibel scale is a log scale which means that for every 10 dB increase there is a 10-fold increase in intensity, hence normal conversation is about 1000 times louder than a whisper. Sounds that are too loud can damage the hearing apparatus and cause severe hearing loss. The pain threshold for sound is about 120 dB. The table below illustrates types of sounds associated with the decibel level and how long the unprotected ear can withstand the sound without being damaged.

Decibel Level	Type	Maximum time until damage without protection
0	Quietest sound you can hear	n/a
30	Whisper	n/a
60	Normal conversations	n/a
90	Lawnmower, shop tools	8 hours per day
100	Chainsaw, snowmobile	2 hours per day
115	Sandblasting, rock concert	15 minutes per day

140	Gun blast, jet engine	0 minutes per day



Structures of the Inner Ear. *BYU-Idaho image created by Isaak Fall 2015*

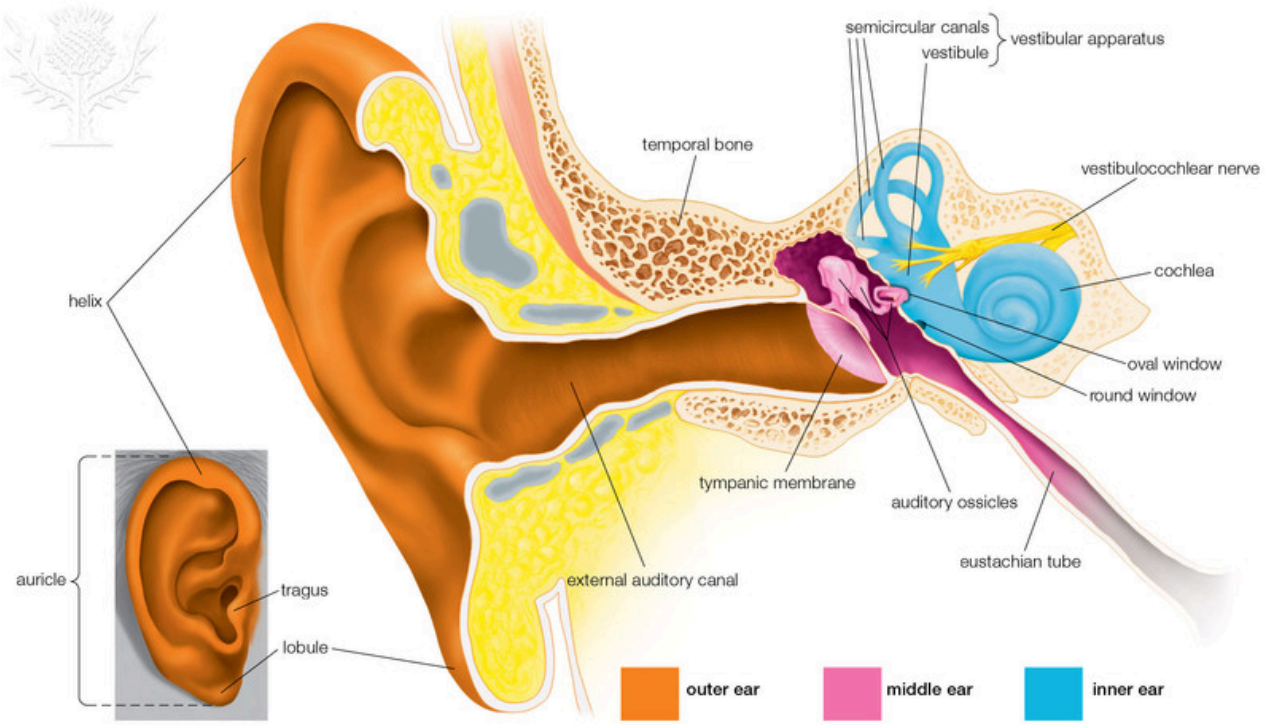
12.4.2

The Hearing Apparatus

Anatomically the ear can be divided into three regions: the **external ear**, the **middle ear**, and the **inner ear**. As the structure of the ear is described try to follow along on the image below. The external ear consists of the **auricle**, or pinna and the **external auditory canal**. The auricle is designed to capture the sound waves and channel them into the auditory canal, which then conducts the sound waves to the **tympanic membrane** (ear drum). Within the auditory canal are ceruminous glands that secrete cerumen (ear wax). Cerumen acts as a lubricant, preventing the auditory canal from drying out, and it has some antibacterial properties to help prevent microorganism from growing in the ear. Additionally, along with the hairs that grow in the auditory canals, the cerumen helps prevent foreign objects from entering the ear.

The tympanic membrane is the boundary between the external ear and the middle ear. Its primary function is to vibrate in response to the sound waves entering the ear. It is composed of a sheet of connective tissue covered on the outside by simple squamous epithelium and on the inside by simple cuboidal epithelium. The tympanic membrane is very sensitive and merely touching it with the end of a Q-tip can elicit sharp pain.

The middle ear is an air filled cavity between the outer ear and the inner ear. The most conspicuous components of the middle ear are the three **auditory ossicles** that form a bridge between the tympanic membrane and the oval window of the inner ear. These three bones are the **malleus** (hammer), the **incus** (anvil), and the **stapes** (stirrup). The malleus sits on the tympanic membrane and the stapes connects to the oval window while the incus sits between the two. The middle ear is connected to the back of the **pharynx** (back of the throat) by the **Eustachian tube**. The Eustachian tube allows the middle ear to equilibrate with the atmospheric pressure. If the pressure in the middle ear is different from atmospheric pressure it will either push or pull on the tympanic membrane resulting in discomfort or pain. The Eustachian tube is typically closed but can open briefly in response to yawning, swallowing or chewing. This is why chewing gum can help alleviate the pressure changes you feel in the ear when you drive over a mountain pass.
The inner ear is imbedded in the petrous portion of the temporal bone and consists of three structures, the **vestibular apparatus**, the **semicircular canals**, and the **cochlea**. It is composed of a network of tunnels in the bone collectively referred to as the **bony labyrinth**. Within the bony labyrinth are membranes that essentially line the tunnels. These membranes form the **membranous labyrinth** and contain the structures for generating action potentials. The vestibular apparatus and semicircular canals are involved with equilibrium and balance while the cochlea is involved with hearing.

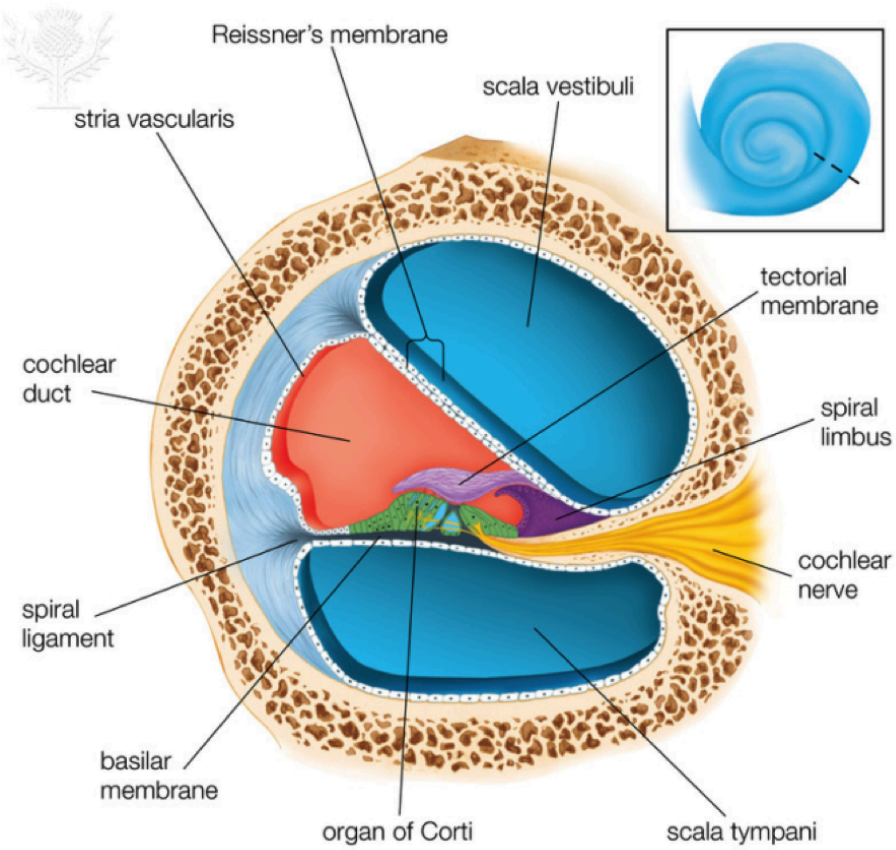


Ear Canal Anatomy. *© 2013 Encyclopædia Britannica, Inc. Downloaded from Image Quest Britannica; BYU-Idaho.*

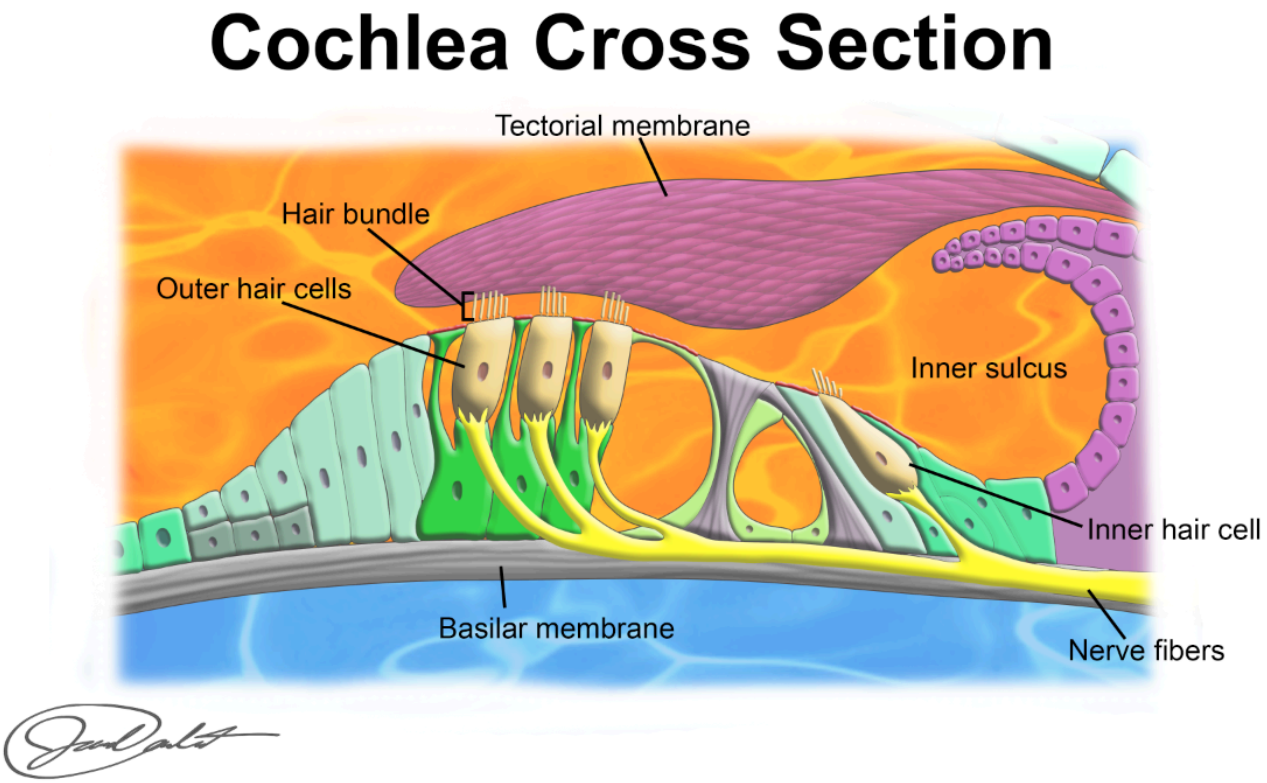
The Cochlea

Before we discuss how sound waves are converted to action potentials, we need to understand the structure of the **cochlea**. This structure gets its name from its shape, cochlea means spiral, or snail shell. The cochlea is a spiral-shaped structure about 3.5 cm long (1.5 inches) that makes 2 ½ turns from top to bottom. It is composed of three parallel chambers that are filled with fluid. The oval window (recall this is a membrane attached to the stapes) communicates with the first chamber, the **scala vestibuli**, which runs the entire length of the cochlea. When the stapes vibrates it causes the fluids in the scala vestibuli to vibrate. At the very tip of the cochlea, the **helicotrema**, the scala vestibuli makes a U-turn and becomes the **scala tympani**. Although they have different names, they are actually one long chamber that folds back on itself. The scala tympani runs parallel to the scala vestibuli and ends at the **round window**. The round window is a thin membrane between the scala tympani and the middle ear. Thus, when the oval window is pushed in by the stapes, the round window bulges out and when the oval window is pulled out, the round window moves in. It is, therefore, acting as a pressure release valve, allowing the fluids in these chambers to vibrate (recall that fluids do not compress). The scala vestibuli and scala tympani are filled with **perilymph**, a fluid that is similar to extracellular fluids. Between these two chambers and also running the length of the cochlea is the **cochlear duct**. This chamber is filled with **endolymph**, which unlike the perilymph, resembles intracellular fluid in composition, and thus has a high K^+ concentration. Within the cochlear duct is the organ that converts mechanical vibrations to electrical action potentials. This structure is the **Organ of Corti** or **Spiral Organ** (see the images below for a cross section of the cochlea and a close up of the spiral organ). The spiral organ sits on the membrane that separates the cochlear duct from the scala tympani, the **basilar membrane**. As we will explain later, this membrane is responsible for detecting sound waves of different frequencies. Structurally, it is narrow and stiff near the oval window and as it moves toward the helicotrema it becomes wider and more limber. This allows each segment to vibrate at a different frequency. Think of the xylophone you had as a child. The keys on one end were very short and when you struck them they emitted a high pitched sound while the keys at the opposite end were long and emitted a low pitched sound when struck, this is basically the structure of the basilar membrane. Separating the cochlear duct from the scala vestibuli is the **vestibular** or **Reisner's membrane**. This is a very flexible membrane that allows the fluid in the cochlear duct to vibrate with the fluid in the scala vestibuli. Located on top of the basilar membrane are four rows of **hair cells**. There are three **outer rows** of hair cells and one **inner row**. These rows run parallel to each other and stretch from the oval window to the helicotrema. As

explained later, these are the receptor cells that will generate action potentials. These cells get their name from the rows of stereocilia on their apical surface. Stereocilia are actually not cilia but instead are more like microvilli. Recall that cilia contain parallel rows of microtubules and are capable of movement whereas microtubules are finger-like projections of the plasma membrane that are supported by microfilaments. In reality, hair cells do have one true cilium called the kinocilium which is adjacent to the longest microtubule. Interestingly, in mammalian cochlea, the kinocilium disappear shortly after birth and no one knows what their function is. Each hair cell has 50-150 stereocilia of different lengths. They are arranged much like the reception bars on your cell phone, gradually increasing in length from one side of the cell to the other. At the point where the stereocilium attaches to the rest of the cell, its diameter is much smaller, creating a hinge-like structure that allows it to bend back and forth. Located on these stereocilia are the mechanically-gated ion channels that will respond to the vibrations of the basilar membrane. Just above the hair cells is another structure called the **tectorial membrane**. It extends like a shelf over the hair cells and the longest stereocilia in the outer three rows of hair cells are embedded in this membrane.



Cross sectional diagram of the cochlea illustrating the three chambers and the organ of Corti. *© 2013 Encyclopædia Britannica, Inc. Downloaded from Image Quest Britannica; BYU-Idaho.*



Organ of Corti. *Produced by BYU-Idaho student Jared Cardinet Fall 2014*

Diagram of the organ of corti and the associated hair cells and neurons.

12.4.3

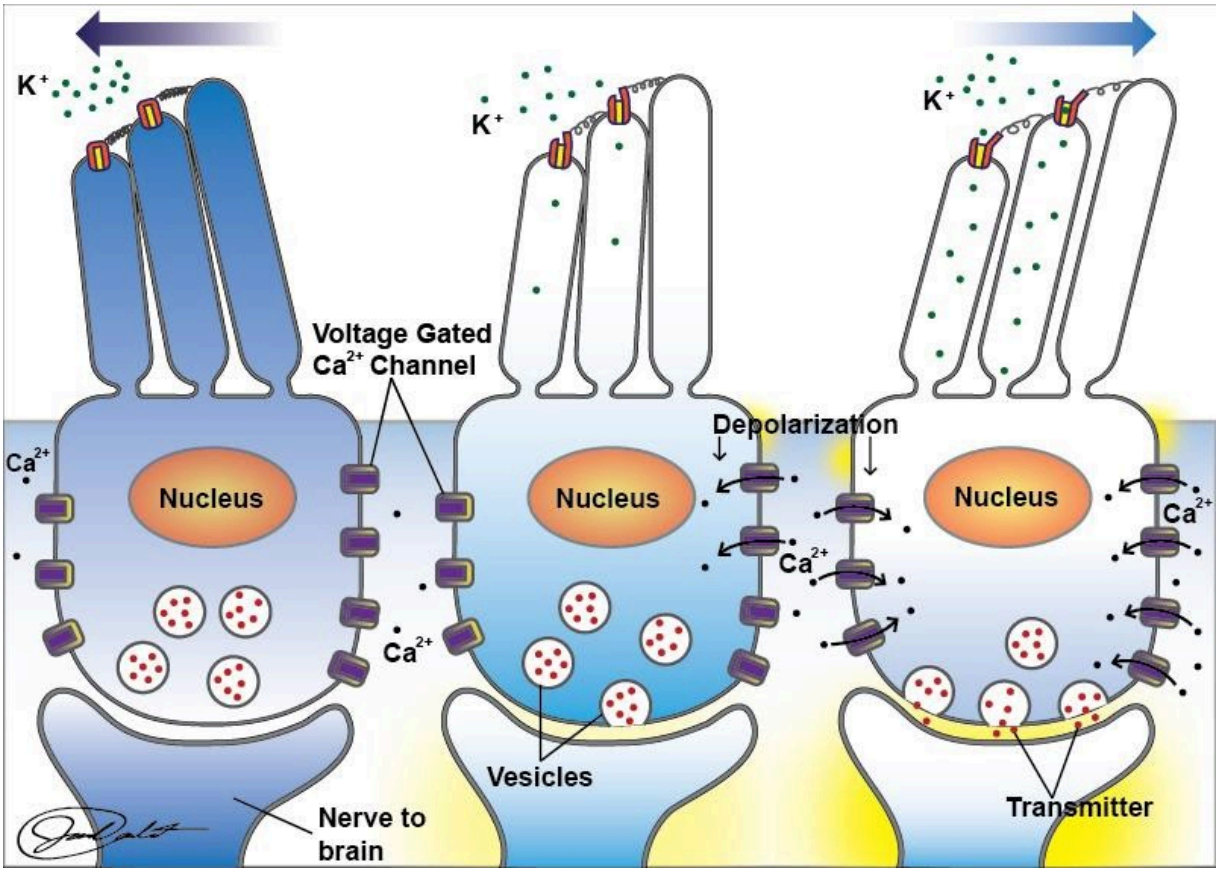
Sound Vibrations to Action Potentials

Transfer of Vibrations in Air to Vibrations in Fluids: The first challenge that our ears face is transferring the vibrations in the air, to vibrations in a fluid. Because the density of the fluid in the inner ear is much greater than the density of air it requires more energy to generate sound waves in the fluid than in the air. Think of being underwater at a swimming pool and listening to people talk, it is very hard to hear and understand. It is the middle ear's responsibility to amplify the sound waves so that their energy is not lost. This is accomplished in two ways. First, the arrangement of the ear ossicles amplifies the sound. Second, and probably more importantly, the tympanic membrane has about 20 times more surface area than the oval window. This size difference results in concentrating the energy on the oval window. Think of how you might move a large rock with a pry bar. You would place the fulcrum close to the stone to gain the maximal mechanical advantage of the bar. the long end of the bar would be analogous to the tympanic membrane and the short end would be analogous to the oval window. These mechanisms are so effective that very little, if any, energy is lost as it is transferred from air waves in the external ear to fluid waves in the internal ear.

Detection of Sound Waves of Different Frequencies: As explained earlier, sound waves of different frequencies are perceived as different pitches. Therefore, the inner ear needs a way of detecting the different frequencies. The structure in the inner ear tasked with this responsibility is the basilar membrane. Recall the design of the basilar membrane, it is narrow and stiff near the oval window and gradually gets wider and more limber as it progresses toward the helicotrema. Think of the example of the xylophone mentioned earlier. When you strike a key on a xylophone it always sounds the same because it always vibrates at the same frequency. Another analogy might be a guitar string. As you tighten a guitar string making it stiffer, it vibrates at a faster rate and produces a sound of a higher pitch. Also on the guitar as you shorten the string by pressing on a fret with your finger the pitch gets higher. At a given tension and length the guitar string always vibrates at the same rate so we always perceive it as the same pitch. The basilar membrane functions in much the same way. Each segment of the membrane has an innate frequency. If it were a guitar string and you plucked it at a certain point along its length it would always vibrate at the same rate at that point. A different point on the basilar membrane would vibrate at a different rate. When a vibration in the fluid reaches the segment of the basilar membrane that has the same innate frequency, it will cause the basilar membrane to vibrate. This phenomenon is known as resonance. Based on this principle of resonance the basilar membrane is able to respond to all of the different frequencies in the sounds we hear, within the range of human hearing.

Conversion of a Sound Wave to an Action Potential: The function of any sensory organ is to convert a sensory stimulus to an action potential that can then be transmitted to the brain. In this case, the sensory signal is the sound wave. The responsibility of converting vibrations into action potentials falls upon the inner row of hair cells in the cochlea. Recall that the apical end of the hair cell contains the stereocilia and that they are arranged in order of ascending lengths from one side of the cell to the other. The membranes of the stereocilia contain mechanically gated cation channels. Extending from the gate of the ion channel to the adjacent, taller, stereocilium is a fibrous protein called a tip link (see image below). When the stereocilia bend toward the longest stereocilium the tension in the tip link increases, pulling the gates on the ion channels open, and when they bend in the opposite direction the tension decreases and the gates close. The stereocilia are bathed in the endolymph of the cochlear duct. Endolymph is similar to intracellular fluid and has a high K^+ concentration. When the gates on the cation channels open, K^+ rushes into the cell, depolarizing the membrane. This depolarization opens voltage gated Ca^{2+} channels on the basal membrane of the hair cell allowing Ca^{2+}

to enter. The influx of Ca^{2+} stimulates the release of neurotransmitter by the hair cell triggering an action potential in the neuron that synapses with the hair cell. The axons of these neurons form the cochlear nerve that transmits the action potential to the auditory cortex of the brain. In hair cells at rest, about 10% of the K^+ ion channels are open resulting in a low frequency of action potentials traveling to the brain when it is perfectly quiet. This allows for both an increase in action potential frequency when hair cells bend toward the longest stereocilium and a decrease in frequency of action potentials when the hair cells bend the other way (see image below).



Conversion of Sound Wave to Action Potential. *Produced by BYU-Idaho student Jared Cardinet Fall 2014*

Hair Cells of the Spiral Organ

Perception of Sound: Once the action potential is generated and sent to the brain it is the function of the auditory cortex to convert that action potential into a perception. Each region of the cochlea is hardwired to its own specific region of the auditory cortex. When that particular region of the brain receives input from the ear we perceive the unique pitch associated with that frequency of the sound wave. It's kind of like a piano where each key is like a different segment of the cochlea. That key is linked to a specific string in the piano such that each time the key is struck we hear the same sound. In this case, the strings would be like a specific region in the auditory cortex. Each time an action potential reaches that specific segment of the auditory cortex we perceive the same sound. Therefore, the pitch is determined by the region of the brain that receives input from the cochlea. Loudness, on the other hand, is determined by the number of action potentials that reach the brain. Recall that the loudness of a sound is a function of the amplitude of the sound wave. Sound waves of higher amplitude cause the hair cells to vibrate more vigorously, which would cause more ion channels to open. This would result in a greater depolarization of the hair cell, more Ca^{2+} entry through the voltage-gated ion channels and more neurotransmitter release. The end result is a greater frequency of action potentials going to the auditory cortex, which is perceived as a louder sound. A common misconception is to equate the frequency of action potentials with the frequency of the sound waves. The frequency of action potentials is a function of the amplitude of the sound wave whereas the frequency of the sound waves determines which portion of the auditory cortex receives the action potentials.

Other Factors Influencing Our Perception of Sound: There are several other factors that impact what we hear. An important consideration is the function of the three outer rows of hair cells. About 90% of the neurons of the cochlear nerve arise from the inner row of hair cells and these are thought to be key to communicating with the auditory cortex. The outer hair cells, on the other hand, are implicated in a process called cochlear amplification. These hair cells have special proteins in their plasma membranes that allow the cell to actively lengthen and shorten. This action can either enhance or reduce the movement of specific regions in the cochlea. The outcome of this action is thought to help focus the sound to specific regions of the spiral organ so that we can better detect the different frequencies of sound waves. Also, recall that only the stereocilia of the outer three rows of hair cells are embedded in the tectorial membrane, those of the inner hair cells are not. What then causes the stereocilia in the inner hair cells to bend? The structure of the spiral organ allows the basilar membrane and the tectorial membrane to function like a bellows. The two membranes form a pocket, the inner sulcus, behind the inner row of hair cells. When the basilar membrane moves up it squeezes the bellows and endolymph flows out of the inner sulcus bending the stereocilia one way and when the basilar membrane moves down the bellows enlarges pulling endolymph into the inner sulcus bending the stereocilia the opposite way.

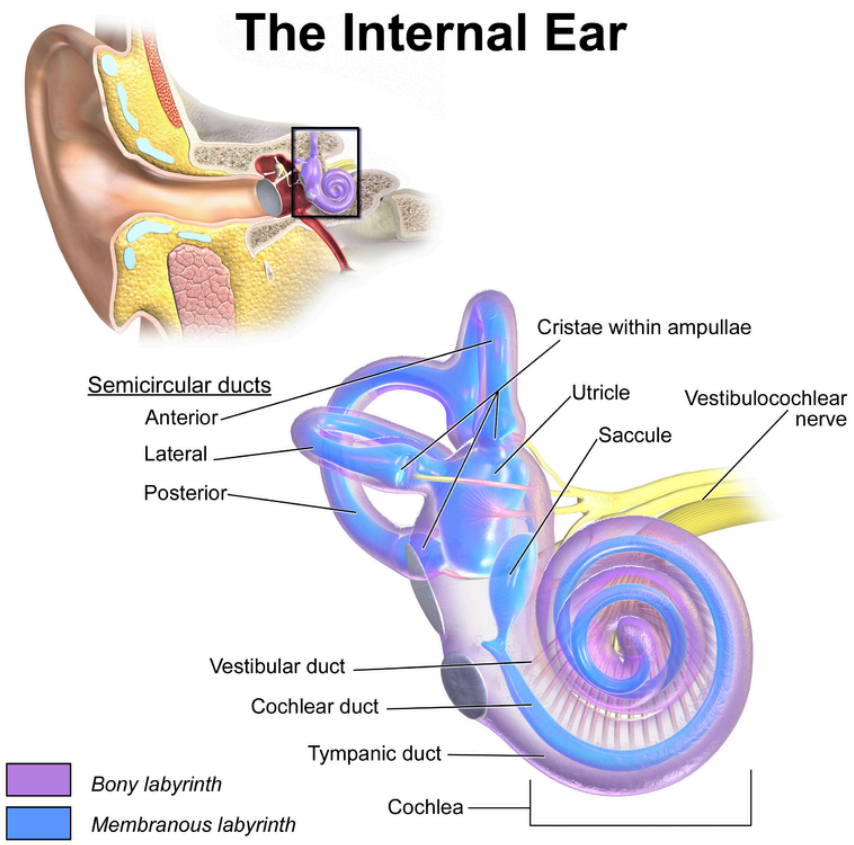
Hearing Loss

There are three forms of hearing loss: conductive, central, and sensorineural. Conductive hearing loss is a result of sound waves being unable to move from the external ear to the inner ear. This can be caused by a plugged ear canal (excessive ear wax), infection of the middle ear or calcification of the stapes to the oval window. Anything that prevents conduction of sound through the external ear or proper vibration of the middle ear bones is termed conductive hearing loss. Central hearing loss results from damage to the auditory cortex, usually caused by a stroke. Sensorineural hearing loss is caused by damage to the structures of the inner ear (hair cells, cochlear neurons, viscous fluid). A common cause of sensorineural hearing loss is exposure to loud sounds. In humans, this loss is irreversible at present. Interestingly, birds have the ability to regenerate hair cells after complete destruction. Study of hair cells in birds may one day lead to the ability to replace damaged hair cells in humans.

It is interesting to note that as we age, we tend to lose the ability to hear very high frequency sounds. This appears to be from a lifetime of wear and tear to the hair cells located closer to the oval window end of the cochlea. This is where the basilar membrane length is shorter and where high energy / high pitch sound waves are experienced.

12.4.4

The Sense of Balance and Equilibrium



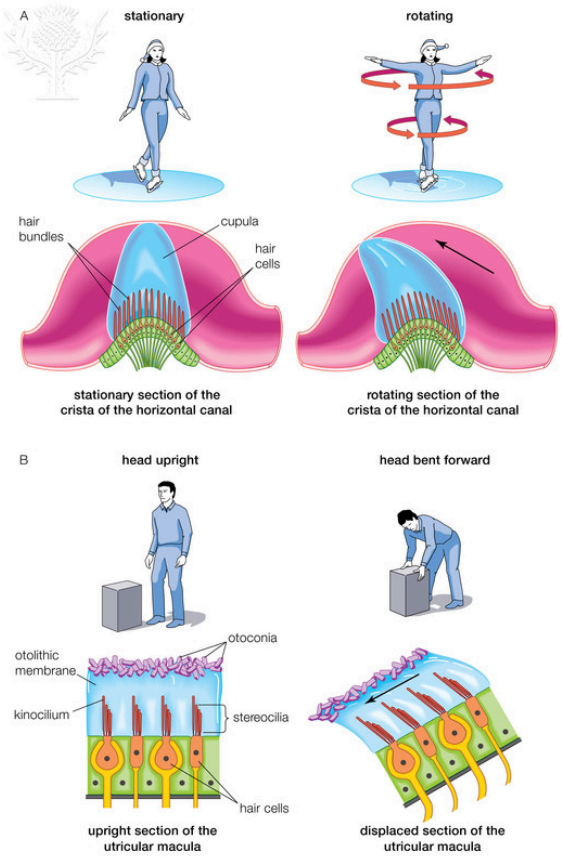
Internal Ear Anatomy. *Title: File:Blausen 0329 EarAnatomy InternalEar.png; Author: Blausen.com staff. "Blausen gallery 2014". Wikiversity Journal of Medicine. DOI:10.15347/wjm/2014.010. ISSN 20018762.;*
Site: https://books.byui.edu/-DnB;
License: This file is licensed under the Creative Commons Attribution 3.0 Unported license.

The senses of equilibrium and balance are very similar to that of hearing in that it involves the use of hair cells and fluid movements. The function of the hair cells involved with equilibrium and balance is the same as that involved with hearing, the difference is in how the stereocilia get bent. Instead of the cochlea, the hair cells of equilibrium are located in two structures: the **vestibular apparatus** and the **semicircular canals**. The vestibular apparatus is designed to sense linear movement as well as the position of our head in relation to gravity. It is sometimes referred to as the static labyrinth. The semicircular canals are designed to detect angular movement and are referred to as the kinetic labyrinth.

The sensory organs in the vestibular apparatus are the **otolith organs**. Within these organs is a specialized sensory structure called the **macula**. It consists of a sheet of hair cells whose stereocilia are imbedded in a gelatinous mass

(see the image below). Within this mass are protein and calcium carbonate crystals known as **otoliths** (ear rocks) that provide weight to the gelatinous mass. It is designed so that when we tilt our heads this mass slides, bending the hair cells. Likewise, when we accelerate linearly, due to the inertia of the mass it lags behind the movement of the head, again bending the hair cells. When we stop, the mass continues to move due to inertia, bending the hair cells in the opposite direction. Each ear has two otolith organs, one oriented horizontally, the **utricle**, and one oriented vertically, the **saccule**. The saccule is responsible for that feeling you love when the elevator starts and stops moving.

The sensory organs of the semicircular canals are designed to detect angular movement. In each ear there are three semicircular canals oriented at right angles to each other. Think of them as being in the frontal, sagittal, and horizontal planes. These canals attach to the vestibular apparatus and are designed so that fluid will flow within the canal. Therefore, if you shake your head side-to-side as if you were saying "no" the fluids in the horizontal canals will move. When you nod your head up and down like when you say "yes", fluid in the sagittal canal will move. And when you move your head back and forth from shoulder to shoulder fluid in the frontal canal will move. At the base of each canal where it attaches to the vestibular apparatus is an enlargement called the **ampulla**. Within the ampulla is the sensory apparatus, the **crista**. The crista contains hair cells that are embedded in a gelatinous mass called the **cupula**. The cupula extends from the hair cells to the top of the ampulla, acting somewhat like a small sail. When the fluid moves through the semicircular canals it pushes the cupula over, which causes the stereocilia on the hair cells to bend (see the image below). Because we have canals oriented in all three planes, any angular movement can be detected. The combination of input from the vestibular apparatus and the semicircular canals informs us about our orientation with respect to gravity as well as any linear or angular movement of our body. Sickness or damage to these organs results in dizziness and even the inability to maintain balance.



Detecting Movement with Crista and Utricular Macula.