Introductory Animal Physiology

Introductory Animal Physiology

Introduction to diversity of structure and function in animals at the tissue and organ system level

SANJA HINIC-FRLOG

Contents

Introduction

Black-capped chickadee in the spring in Erindale Park, Mississauga. This photo and textbook cover photo are taken by Margeaux Gaerlan, Clyde Tan, and Nadine Le.

This book is an introduction to the diversity of structure and function in animals at the tissue and organ system level. The focus of this book is on principles and mechanisms that sustain life and maintain homeostasis, including water balance, gas exchange, acquisition and transport of oxygen and nutrients, temperature regulation, electrical and chemical signal transmission, sensory processing, and locomotion. The content in this open textbook was adapted from other open textbooks (cc by 4.0) resources or created/written by Sanja Hinic-Frlog and collaborators. Collaborators include: Jessica Hanley, Simone Laughton, and invited undergraduate and graduate students at the University of Toronto Mississauga.

Thank You

Special thanks to our students Arlene Cortiguoerra, Kim Nguyen, and Abijah Nwakanma for organizing the adapted content and creating new resources and images for this open textbook.

Acknowledgements: eCampusOntario

Share

If you adopt this book, as a core or supplemental resource, please report your adoption in order for us to celebrate your support of students' savings. Report your commitment at www.openlibrary.ecampusontario.ca.

We invite you to adapt this book further to meet your and your students' needs. Please let us know if you do! If you would like to use Pressbooks, the platform used to make this book, contact eCampusOntario for an account using open@ecampusontario.ca.

If this text does not meet your needs, please check out our full library at www.openlibrary.ecampusontario.ca. If you still cannot find what you are looking for, connect with colleagues and eCampusOntario to explore creating your own open education resource (OER).

About eCampusOntario

eCampusOntario is a not-for-profit corporation funded by the Government of Ontario. It serves as a centre of excellence in online and technology-enabled learning for all publicly funded colleges and universities in Ontario and has embarked on a bold mission to widen access to post-secondary education and training in Ontario. This textbook is part of eCampusOntario's open textbook library, which provides free learning resources in a wide range of subject areas. These open textbooks can be assigned by instructors for their classes and can be downloaded by learners to electronic devices or printed for a low cost by our printing partner, The University of Waterloo. These free and open educational resources are customizable to meet a wide range of learning needs, and we invite instructors to review and adopt the resources for use in their courses.

Copyright – Introductory Animal Physiology

Icons Legend

Legend of Icons that are used in this eTextbook.

Icon Title	Icon Meaning Description	Icon Image
Challenge / Game	An opportunity to challenge your understanding or to engage in a game related to a concept or topic.	
Complete Tasks	Tasks or activities to be completed to consolidate understanding and skills.	
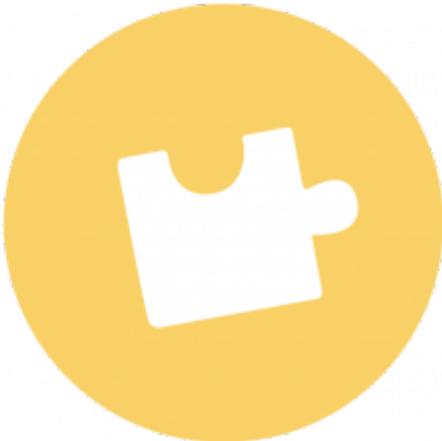
| Discussion | An opportunity to discuss a concept or a topic with your peers. | |

| Explore Further Resources | Additional resources that can be used to further explore a concept or topic. | |

Listen to Podcast

Additional articles, studies, etc. that can be read to learn more about a concept or topic.

Important

Items that you should take note of that are useful to help with understanding a particular concept or topic.

Learning Outcomes

What you should know or be able to do after reading a chapter and having engaged in the related activities for that chapter.

Peer Collaboration

Activities that you can work on with your peers.

Presentation | A link to information from a presentation or in presentation format.
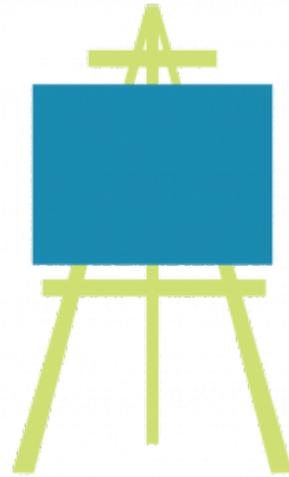
Questions | Questions that can be used to self-check your understanding or skill level.

Quiz

An opportunity to test your skills and understanding.

Research

A link to research on a concept or a topic or items that may require further research.
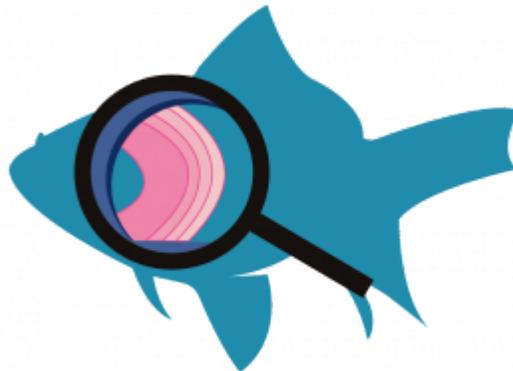
Suggestions

Items that are suggestions for further thought or study.

Think

Items that may require additional thought and consideration.

Watch Video

A video resource that provides information related to a concept or topic being discussed.

CHAPTER 1. INTRODUCTION TO PHYSIOLOGY

1.1. Define physiology and explain its importance and connections to other fields of biological studies;

1.2. Describe physical, chemical and evolutionary limitations to sustaining life; and,

1.3. Provide a general description of and some examples of homeostasis.

1.1 – What is Physiology?

The content of this chapter was adapted from the *Concepts of Biology-1st Canadian Edition* open textbook by Charles Molnar and Jane Gair ([Chapter 11 – Introduction to the body's systems](#)).

1.1. Define physiology and explain its importance and connections to other fields of biological studies.

Physiology is a study of how animal form or structure and function sustain life and shape responses to environmental conditions. The arctic fox (Figure 1.1), a complex animal that has adapted to its environment, illustrates the relationships between an animal's form and function. The multicellular bodies of animals consist of tissues that make up more complex organs and organ systems. The organ systems of an animal maintain homeostasis within the multicellular body. These systems are adapted to obtain the necessary nutrients and other resources needed by the cells of the body, to remove the wastes those cells produce, to coordinate the activities of the cells, tissues, and organs throughout the body, and to coordinate the many responses of the individual organism to its environment.

Figure 1.1 Arctic fox in Wapusk National Park, Manitoba, Canada. Photo by By Ansgar Walk – photo taken by Ansgar Walk, CC BY 2.5, https://commons.wikimedia.org/w/index.php?curid=905259

Some examples of arctic fox physiological adaptions to its environment include:

- Fur with insulative properties to keep the arctic fox warm;
- Reduced heat loss through adaptations to body components such as shorter ears, legs, muzzle;
- Ability to stand on cold surfaces due to capillary rete in pads;
- Can reduce metabolic rate to conserve energy, especially helpful in times of food scarcity.

SOURCE: PESTRUD, P. (1991). *Adaptations by the Arctic Fox (Alopex lagopus) to the Polar Winter.* Arctic, 44(2), 132-138. Retrieved 2018 February 5 from http://www.jstor.org/stable/40511073.

The study of physiology is interrelated with other fields of biological studies, such as

- Anatomy (how organisms are structured),
- Biogeography (spatial and temporal distribution of organisms),
- Biomechanics (the study of how organisms move),
- Conservation Biology (the study of the natural environment and ecosystems),
- Ecology (the study of how organisms interact with other living organisms and their environment),
- Ethology (the study of animal behaviour), and
- Other fields of Biology.

Question 1.1

Define physiology in your own words.

Watch this video to see how Arctic foxes find prey in the winter.

1.2 – Limits of Life

This section was adapted from Shulze-Makuch, Shulze-Makuch, and Houtkooper, 2015. The physical, chemical, and physiological limits of life. Life 5(3): 1472-1486. doi:10.3390/life5031472.

1.2. Describe physical, chemical and evolutionary limitations to sustaining life.

Life on Earth exists in nearly every niche on our planet. However, life forms are not equally successful in all conditions on Earth, and one challenge is to find under which conditions life can be sustained. Certainly, life on Earth is very adaptable, which has led to an immense biomass and an incredible biodiversity. Therefore, it is a challenge to find where the limits to this adaptability are.

Question 1.2

Try to come up with the best answer for the following question:
Which of the following may be the most limiting physical factor for animals living in aquatic environments? Why?
a. locomotion
b. being wet all the time
c. access to nutrients
d. access to oxygen

On Earth, life is based on carbon as a major building block, water as a solvent and chemical bonds and light as life-sustaining energy sources, which seems to be an ideal combination on a terrestrial planet like Earth, which has an average surface temperature of about 15 °C under a pressure of 1 bar. In principle, other building blocks than carbon are feasible, such as silicon, but these would require very different planetary conditions to be used in a comparable fashion to carbon on Earth. Water as a solvent has great benefits and also some challenges (reviewed by Schulze-Makuch and Irwin [Schulze-Makuch & Irwin., 2008]), but being the most abundant molecule on our planet that exists in liquid form, life had simply to adapt to some of its drawbacks. Light is plentiful on our planet, and organic compounds with covalent bonds are versatile at average Earth temperatures, thus providing a powerful combination that resulted on Earth in a biosphere with a large biomass and an incredible biodiversity, including complex life.

There are limits to the conditions under which life can exist on our planet. Most pronounced is the temperature envelope under which active life can exist. Organismic growth can usually occur at temperatures from at least –15 °C to about 113 °C. There are also reports in the literature that

the temperature range may even be broader. For example, metabolic activity was inferred down to temperatures of –40 °C due to anomalous concentrations of gases [Campen et al., 2003], and the upper temperature limit may be as high as about 122 °C. At that temperature, methanogenic archaea could be cultured under a pressure of 20 MPa [Takai et al., 2008], only limited by the solubility of lipids in water and protein stability [McKay, 2014]. In principle, if the biochemistry of organisms could be adapted to these extremes, perhaps even higher temperatures may be tolerated, but the practical limit due to energetic and biochemical constraints under which life can still metabolize and reproduce is surely much lower [Schulze-Makuch, 2015]. Hyperthermophilic microorganisms require specialized cell components, like proteins and membranes, to be stable and function at high temperatures. Particularly, at temperatures of 100 °C and beyond, some low molecular weight compounds, such as ATP and NAD, hydrolyze quite rapidly, and thermolabile amino acids, like cysteine and glutamic acid, are decomposing [Stetter, 1999]. The pressure tolerance of life, though, is high and extends to at least 1100 bar [1] [Stan-Lotter, 2007].

Organisms, particularly microbes and other microscopic organisms, are quite tolerant to extreme pH values, from just below 0 to about 13. Ferroplasma sp. and Cephalosporium are examples of organisms that live at low pH-values, Natronobacterium, and several species of protists and rotifers are examples of organisms that live at very high pH-values [Schulze-Makuch & Irwin, 2008, Baross et al., 2007].

However, life on Earth is relatively sensitive to a lack of water: bacteria, archaea, and fungi can only metabolize at water activities down to about 0.6 [Stevenson, et al., 2014]. Adaptation to water with high salt content, however, is quite common, as some Halobacteria and archaea can grow in 35% NaCl solution [Schulze-Makuch & Irwin, 2008].

Another physical limit to life is radiation, both UV, and ionizing radiation. Tolerances to radiation vary widely. Tardigrades, microscopic animals that usually live in mosses and lichen, can withstand ionizing radiation doses up to 5000 Gy when in the dormant state [Schulze-Makuchh & Seckbach, 2013] and display additional special adaptation traits, such as anhydrobiosis and cryptobiosis [Watanabe, 2006]. Deinococcus radiodurans can still tolerate higher radiation doses and grow at doses upward of 10,000 Gy.

One non-chemical limitation of life on earth is related to scaling of body size. You can find more details about scaling in animals in this article.

---

1. Stan-Lotter, H. Extremophiles, the physicochemical limits of life (growth and survival). In Complete Course in Astrobiology; Horneck, G., Rettberg, P., Eds.; Wiley-VCH: Weinheim, Germany, 2007; pp. 121–150.

Question 1.3

Some examples of physical and chemical limitations of life were listed above. One aspect that was not discussed is size. How do you think size is a physical limitation for sustaining life for animals? Find one specific example with one of your peers. Once you have found your example, you should think about why the size is the limit in your example.

Here is a link to help you think about the physical and chemical limitations of life, specifically animal size. This is a [podcast](#) on how climate affected size changes of beetles in Canada.

Question 1.4

Describe five different limitations to life on Earth.

1.3 – Homeostasis

The content of this chapter was adapted from the *Concepts of Biology-1st Canadian Edition* open textbook by Charles Molnar and Jane Gair ([Chapter 11.1 – Homeostasis and Osmoregulation](#)) and *Anatomy and Physiology* open textbook ([Chapter 1.5 – Homeostasis](#)).

1.3. Provide a general description of and some examples of homeostasis.

In order to function properly, cells require appropriate conditions such as proper temperature, pH, and appropriate concentration of diverse chemicals. These conditions may, however, change from one moment to the next. Organisms are able to maintain internal conditions within a narrow range almost constantly, despite environmental changes, through homeostasis (literally, "steady state"). For example, an organism needs to regulate body temperature through the thermoregulation process. Organisms that live in cold climates, such as the polar bear, have body structures that help them withstand low temperatures and conserve body heat. Structures that aid in this type of insulation include fur, feathers, blubber, and fat. In hot climates, organisms have methods (such as perspiration in humans or panting in dogs) that help them to shed excess body heat.

Homeostasis refers to the relatively stable state inside the body of an animal. Animal organs and organ systems constantly adjust to internal and external changes in order to maintain this steady state. Examples of internal conditions maintained homeostatically are the level of blood glucose, body temperature, blood calcium level. These conditions remain stable because of physiologic processes that result in negative feedback relationships. If the blood glucose or calcium rises, this sends a signal to organs responsible for lowering blood glucose or calcium. The signals that restore the normal levels are examples of negative feedback. When homeostatic mechanisms fail, the results can be unfavorable for the animal. Homeostatic mechanisms keep the body in dynamic equilibrium by constantly adjusting to the changes that the body's systems encounter. Even an animal that is apparently inactive is maintaining this homeostatic equilibrium. Two examples of factors that are regulated homeostatically are temperature and water content. The processes that maintain homeostasis of these two factors are called thermoregulation and osmoregulation.

Homeostasis

The goal of homeostasis is the maintenance of equilibrium around a specific value of some aspect of the body or its cells called a set point. While there are normal fluctuations from the set point, the body's systems will usually attempt to go back to this point. A change in the internal or external environment is called a stimulus and is detected by a receptor; the response of the system is to adjust the activities of the system so the value moves back toward the set point. For instance, if the body becomes too warm, adjustments are made to cool the animal. If glucose levels in the blood rise after a meal, adjustments are made to lower them and to get the nutrient into tissues that need it or to store it for later use.

When a change occurs in an animal's environment, an adjustment must be made so that the internal environment of the body and cells remains stable. The receptor that senses the change in the environment is part of a feedback mechanism. The stimulus–temperature, glucose, or calcium levels–is detected by the receptor. The receptor sends information to a control center, often the brain, which relays appropriate signals to an effector organ that is able to cause an appropriate change, either up or down, depending on the information the sensor was sending.

Thermoregulation

Animals can be divided into two groups: those that maintain a constant body temperature in the face of differing environmental temperatures, and those that have a body temperature that is the same as their environment and thus varies with the environmental temperature. Animals that do not have internal control of their body temperature are called ectotherms. The body temperature of these organisms is generally similar to the temperature of the environment, although the individual organisms may do things that keep their bodies slightly below or above the environmental temperature. This can include burrowing underground on a hot day or resting in the sunlight on a cold day. The ectotherms have been called cold-blooded, a term that may not apply to an animal in the desert with a very warm body temperature.

An animal that maintains a constant body temperature in the face of environmental changes is called an endotherm. These animals are able to maintain a level of activity that an ectothermic animal cannot because they generate internal heat that keeps their cellular processes operating optimally even when the environment is cold.

Watch this [Discovery Channel video](#) on thermoregulation to see illustrations of the process in a variety of animals.

Animals conserve or dissipate heat in a variety of ways. Endothermic animals have some form of insulation. They have fur, fat, or feathers. Animals with thick fur or feathers create an insulating layer of air between their skin and internal organs. Polar bears and seals live and swim in a subfreezing environment and yet maintain a constant, warm, body temperature. The arctic fox, for example, uses its fluffy tail as extra insulation when it curls up to sleep in cold weather. Mammals can increase body heat production by shivering, which is an involuntary increase in muscle activity. In addition, arrector pili muscles can contract causing individual hairs to stand up when the individual is cold. This increases the insulating effect of the hair. Humans retain this reaction, which does not have the intended effect on our relatively hairless bodies; it causes "goose bumps" instead. Mammals use layers of fat as insulation also. Loss of significant amounts of body fat will compromise an individual's ability to conserve heat.

Ectotherms and endotherms use their circulatory systems to help maintain body temperature. Vasodilation, the opening up of arteries to the skin by relaxation of their smooth muscles, brings more blood and heat to the body surface, facilitating radiation and evaporative heat loss, cooling the body. Vasoconstriction, the narrowing of blood vessels to the skin by contraction of their smooth muscles, reduces blood flow in peripheral blood vessels, forcing blood toward the core and vital organs, conserving heat. Some animals have adaptions to their circulatory system that enable them to transfer heat from arteries to veins that are flowing next to each other, warming blood returning to the heart. This is called a countercurrent heat exchange; it prevents the cold venous blood from cooling the heart and other internal organs. The countercurrent adaptation is found in dolphins, sharks, bony fish, bees, and hummingbirds.

Some ectothermic animals use changes in their behavior to help regulate body temperature. They simply seek cooler areas during the hottest part of the day in the desert to keep from getting too warm. The same animals may climb onto rocks in the evening to capture heat on a cold desert night before entering their burrows.

Thermoregulation is coordinated by the nervous system ([Figure 1.2](#)). The processes of temperature control are centered in the hypothalamus of the advanced animal brain. The hypothalamus maintains the set point for body temperature through reflexes that cause vasodilation or vasoconstriction and shivering or sweating. The sympathetic nervous system under control of the hypothalamus directs the responses that effect the changes in temperature loss or gain that return the body to the set point. The set point may be adjusted in some instances. During an infection, compounds called

pyrogens are produced and circulate to the hypothalamus resetting the thermostat to a higher value. This allows the body's temperature to increase to a new homeostatic equilibrium point in what is commonly called a fever. The increase in body heat makes the body less optimal for bacterial growth and increases the activities of cells so they are better able to fight the infection.
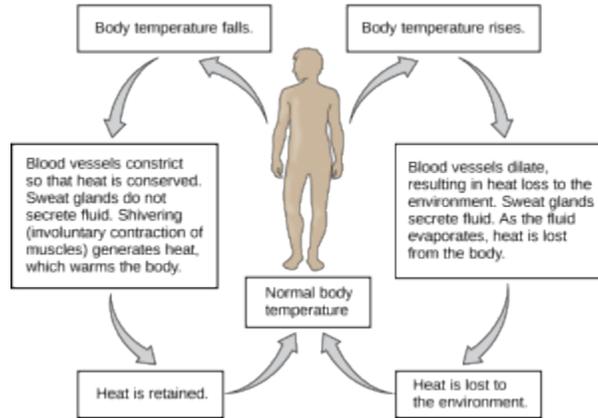
Question 1.5

When bacteria are destroyed by leukocytes, pyrogens are released into the blood. Pyrogens reset the body's thermostat to a higher temperature, resulting in fever. How might pyrogens cause the body temperature to rise?
Note: Pyrogens increase body temperature by causing the blood vessels to constrict, inducing shivering, and stopping sweat glands from secreting fluid

Question 1.6

What is homeostasis?

Question 1.7

Describe a thermoregulatory homeostatic loop.

Question 1.8

Describe an osmoregulatory homeostatic loop.

Examples of maintenance of homeostasis through negative feedback

Negative feedback is a mechanism that reverses a deviation from the set point. Therefore, negative feedback maintains body parameters within their normal range. The maintenance of homeostasis by negative feedback goes on throughout the body at all times, and an understanding of negative feedback is thus fundamental to an understanding of human physiology. A negative feedback system has three basic components (Figure 1.3a). A **sensor**, also referred to a receptor, is a component of a feedback system that monitors a physiological value. This value is reported to the control center. The **control center** is the component in a feedback system that compares the value to the normal range. If the value deviates too much from the set point, then the control center activates an effector. An **effector** is the component in a feedback system that causes a change to reverse the situation and return the value to the normal range.

(a) Negative feedback loop   (b) Body temperature regulation
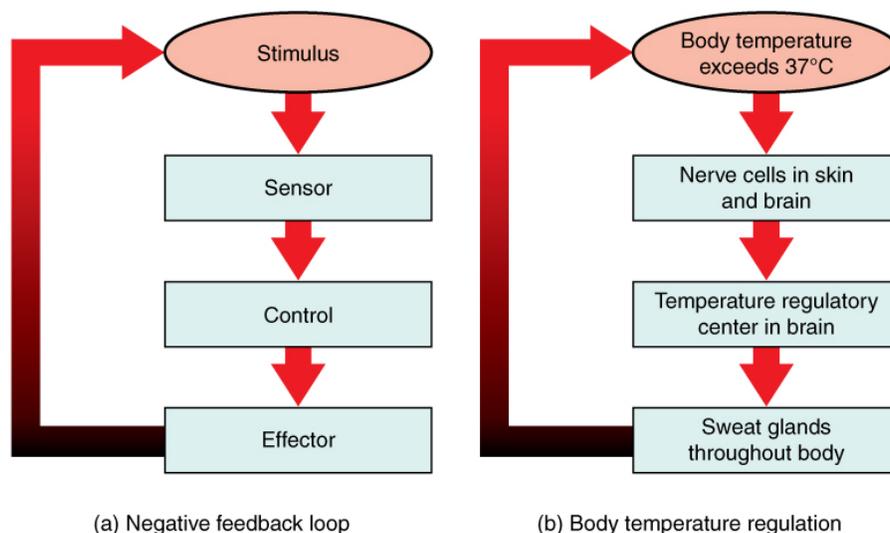
Figure 1.3. Negative feedback loop. In a negative feedback loop, a stimulus—a deviation from a set point—is resisted through a physiological process that returns the body to homeostasis. (a) A negative feedback loop has four basic parts. (b) Body temperature is regulated by negative feedback.

In order to set the system in motion, a stimulus must drive a physiological parameter beyond its normal range (that is, beyond homeostasis). This stimulus is "heard" by a specific sensor. For example, in the control of blood glucose, specific endocrine cells in the pancreas detect excess glucose (the stimulus) in the bloodstream. These pancreatic beta cells respond to the increased level of blood glucose by releasing the hormone insulin into the bloodstream. The insulin signals skeletal muscle fibers, fat cells (adipocytes), and liver cells to take up the excess glucose, removing it from the bloodstream. As glucose concentration in the bloodstream drops, the decrease in concentration—the actual negative feedback—is detected by pancreatic alpha cells, and insulin release stops. This prevents blood sugar levels from continuing to drop below the normal range.

Humans have a similar temperature regulation feedback system that works by promoting either heat loss or heat gain (Figure 1.3b). When the brain's temperature regulation center receives data from the sensors indicating that the body's temperature exceeds its normal range, it stimulates a cluster of brain cells referred to as the "heat-loss center." This stimulation has three major effects:

- Blood vessels in the skin begin to dilate allowing more blood from the body core to flow to the surface of the skin allowing the heat to radiate into the environment.
- As blood flow to the skin increases, sweat glands are activated to increase their output. As the sweat evaporates from the skin surface into the surrounding air, it takes heat with it.
- The depth of respiration increases, and a person may breathe through an open mouth instead of through the nasal passageways. This further increases heat loss from the lungs.

In contrast, activation of the brain's heat-gain center by exposure to cold reduces blood flow to the skin, and blood returning from the limbs is diverted into a network of deep veins. This arrangement traps heat closer to the body core and restricts heat loss. If heat loss is severe, the brain triggers an increase in random signals to skeletal muscles, causing them to contract and producing shivering. The muscle contractions of shivering release heat while using up ATP. The brain triggers the thyroid gland in the endocrine system to release thyroid hormone, which increases metabolic activity and heat production in cells throughout the body. The brain also signals the adrenal glands to release epinephrine (adrenaline), a hormone that causes the breakdown of glycogen into glucose, which can be used as an energy source. The breakdown of glycogen into glucose also results in increased metabolism and heat production.

Watch this video to learn more about water concentration in the body.

Water concentration in the body is critical for proper functioning. A person's body retains very tight

control on water levels without conscious control by the person. Watch this video to learn more about water concentration in the body. Which organ has primary control over the amount of water in the body?

Examples of maintenance of homeostasis through positive feedback

Positive feedback intensifies a change in the body's physiological condition rather than reversing it. A deviation from the normal range results in more change, and the system moves farther away from the normal range. Positive feedback in the body is normal only when there is a definite end point. Childbirth and the body's response to blood loss are two examples of positive feedback loops that are normal but are activated only when needed.

Childbirth at full term is an example of a situation in which the maintenance of the existing body state is not desired. Enormous changes in the mother's body are required to expel the baby at the end of pregnancy. And the events of childbirth, once begun, must progress rapidly to a conclusion or the life of the mother and the baby are at risk. The extreme muscular work of labor and delivery are the result of a positive feedback system (Figure 1.4).
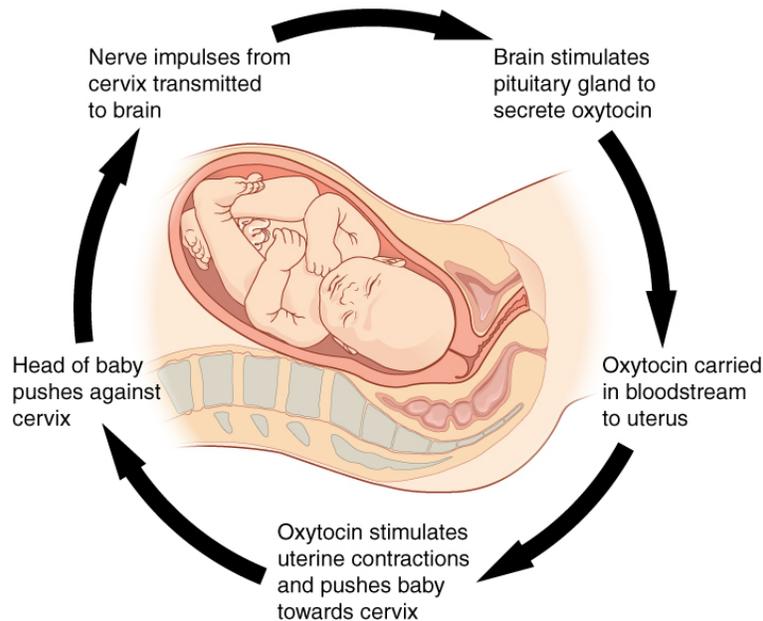
Figure 1.4. Positive feedback loop. Normal childbirth is driven by a positive feedback loop. A positive feedback loop results in a change in the body's status, rather than a return to homeostasis.

The first contractions of labor (the stimulus) push the baby toward the cervix (the lowest part of the uterus). The cervix contains stretch-sensitive nerve cells that monitor the degree of stretching (the

sensors). These nerve cells send messages to the brain, which in turn causes the pituitary gland at the base of the brain to release the hormone oxytocin into the bloodstream. Oxytocin causes stronger contractions of the smooth muscles in the uterus (the effectors), pushing the baby further down the birth canal. This causes even greater stretching of the cervix. The cycle of stretching, oxytocin release, and increasingly more forceful contractions stops only when the baby is born. At this point, the stretching of the cervix halts, stopping the release of oxytocin.

A second example of positive feedback centers on reversing extreme damage to the body. Following a penetrating wound, the most immediate threat is excessive blood loss. Less blood circulating means reduced blood pressure and reduced perfusion (penetration of blood) to the brain and other vital organs. If perfusion is severely reduced, vital organs will shut down and the person will die. The body responds to this potential catastrophe by releasing substances in the injured blood vessel wall that begin the process of blood clotting. As each step of clotting occurs, it stimulates the release of more clotting substances. This accelerates the processes of clotting and sealing off the damaged area. Clotting is contained in a local area based on the tightly controlled availability of clotting proteins. This is an adaptive, life-saving cascade of events.

Question 1.9

After you eat lunch, nerve cells in your stomach respond to the distension (the stimulus) resulting from the food. They relay this information to _____.
a. a control center
b. a set point
c. effectors
d. sensors

Question 1.10

Stimulation of the heat-loss center causes _____.
a. blood vessels in the skin to constrict
b. breathing to become slow and shallow
c. sweat glands to increase their output
d. All of the above.

Question 1.11

Which of the following is an example of a normal physiologic process that uses a positive feedback loop?
a. blood pressure regulation
b. childbirth
c. regulation of fluid balance
d. temperature regulation

Question 1.12

Identify the four components of a negative feedback loop and explain what would happen if secretion of a body chemical controlled by a negative feedback system became too great.

Question 1.13

What regulatory processes would your body use if you were trapped by a blizzard in an unheated, uninsulated cabin in the woods?

CHAPTER 2. GASEOUS EXCHANGE

2.1. Discuss general respiration advantages and disadvantages with respect to aquatic and terrestrial environments;

2.2. Describe select respiratory structures in aquatic and terrestrial animals;

2.3. Explain relationships between respiratory structure and function in generalized aquatic and terrestrial animal systems; and,

2.4. Compare counter- and crosscurrent exchange using specific respiratory system examples.

2.1 – Introduction to Respiratory Systems

The content of this chapter was adapted from the *Concepts of Biology-1st Canadian Edition* open textbook by Charles Molnar and Jane Gair ([Chapter 11 – Introduction to the body's systems](#)).

2.1. Discuss general respiration advantages and disadvantages with respect to aquatic and terrestrial environments.

Breathing is an involuntary event. How often a breath is taken and how much air is inhaled or exhaled are tightly regulated by the respiratory center in the brain. Humans, when they aren't exerting themselves, breathe approximately 15 times per minute on average. Canines, like the dog in [Figure 2.1](#), have a respiratory rate of about 15–30 breaths per minute. With every inhalation, air fills the lungs, and with every exhalation, air rushes back out. That air is doing more than just inflating and deflating the lungs in the chest cavity. The air contains oxygen that crosses the lung tissue, enters the bloodstream, and travels to organs and tissues. Oxygen (O_2) enters the cells where it is used for metabolic reactions that produce ATP, a high-energy compound. At the same time, these reactions release carbon dioxide (CO_2) as a by-product. CO_2 is toxic and must be eliminated. Carbon dioxide exits the cells, enters the bloodstream, travels back to the lungs, and is expired out of the body during exhalation.

Figure 2.1. Lungs, which appear as nearly transparent tissue surrounding the heart in this X-ray of a dog (left), are the central organs of the respiratory system. The left lung is smaller than the right lung to accommodate space for the heart. A dog's nose (right) has a slit on the side of each nostril. When tracking a scent, the slits open, blocking the front of the nostrils. This allows the dog to exhale through the now-open area on the side of the nostrils without losing the scent that is being followed. (credit a: modification of work by Geoff Stearns; credit b: modification of work by Cory Zanker)

The primary function of the respiratory system is to deliver oxygen to the cells of the body's tissues and remove carbon dioxide, a cell waste product. The main structures of the human respiratory system are the nasal cavity, the trachea, and the lungs.

All aerobic organisms require oxygen to carry out their metabolic functions. Along the evolutionary tree, different organisms have devised different means of obtaining oxygen from the surrounding atmosphere. The environment in which the animal lives greatly determines how an animal respires. The complexity of the respiratory system is correlated with the size of the organism. As animal size increases, diffusion distances increase and the ratio of surface area to volume drops. In unicellular organisms, diffusion across the cell membrane is sufficient for supplying oxygen to the cell ([Figure 2.2](#)). Diffusion is a slow, passive transport process. In order for diffusion to be a feasible means of providing oxygen to the cell, the rate of oxygen uptake must match the rate of diffusion across the membrane. In other words, if the cell were very large or thick, diffusion would not be able to provide oxygen quickly enough to the inside of the cell. Therefore, dependence on diffusion as a means of obtaining oxygen and removing carbon dioxide remains feasible only for small organisms or those with highly-flattened bodies, such as many flatworms (Platyhelminthes). Larger organisms had to evolve specialized respiratory tissues, such as gills, lungs, and respiratory passages accompanied by a complex circulatory system, to transport oxygen throughout their entire body.

Figure 2.2. The cell of the unicellular algae Ventricaria ventricosa is one of the largest known, reaching one to five centimeters in diameter. Like all single-celled organisms, V. ventricosa exchanges gases across the cell membrane.

Direct diffusion

For small multicellular organisms, diffusion across the outer membrane is sufficient to meet their oxygen needs. Gas exchange by direct diffusion across surface membranes is efficient for organisms less than 1 mm in diameter. In simple organisms, such as cnidarians and flatworms, every cell in the body is close to the external environment. Their cells are kept moist and gases diffuse quickly via direct diffusion. Flatworms are small, literally flatworms, which 'breathe' through diffusion across the outer membrane (Figure 2.3). The flat shape of these organisms increases the surface area for diffusion, ensuring that each cell within the body is close to the outer membrane surface and has access to oxygen. If the flatworm had a cylindrical body, then the cells in the center would not be able to get oxygen.

Figure 2.3. This flatworm's process of respiration works by diffusion across the outer membrane. (credit: Stephen Childs)

Earthworms and amphibians use their skin (integument) as a respiratory organ. A dense network of capillaries lies just below the skin and facilitates gas exchange between the external environment and the circulatory system. The respiratory surface must be kept moist in order for the gases to dissolve and diffuse across cell membranes. Organisms that live in water need to obtain oxygen from the water. Oxygen dissolves in water but at a lower concentration than in the atmosphere. The atmosphere has roughly 21 percent oxygen. In water, the oxygen concentration is much smaller than that.

Question 2.1

Which physical and chemical factors affect the rate of diffusion of a gas and how? (You will be able to find some answers to this question in the next section of this chapter).

Question 2.2

How are some fish able to breathe on land?

2.2 – Respiratory Structures and Their Function

The content of this chapter was adapted from the *Concepts of Biology-1st Canadian Edition* open textbook by Charles Molnar and Jane Gair ([Chapter 20 -The respiratory system](#)).

2.2. Describe select respiratory structures in aquatic and terrestrial animals.

2.3. Explain relationships between respiratory structure and function in generalized aquatic and terrestrial animal systems.

Ventilation and Perfusion

Two important aspects of gas exchange in the lung are ventilation and perfusion. Ventilation is the movement of air into and out of the lungs, and perfusion is the flow of blood in the pulmonary capillaries. For gas exchange to be efficient, the volumes involved in ventilation and perfusion should be compatible. Therefore, ventilatory structures and perfusion surfaces have evolved to maximize this compatibility in both aquatic and terrestrial environment and examples of such structures and surfaces will be provided next.

Skin and gills

Fish and many other aquatic organisms have evolved gills to take up the dissolved oxygen from water ([Figure 2.4](#)). Gills are thin tissue filaments that are highly branched and folded. When water passes over the gills, the dissolved oxygen in water rapidly diffuses across the gills into the bloodstream. The circulatory system can then carry the oxygenated blood to the other parts of the body. In animals that contain coelomic fluid instead of blood, oxygen diffuses across the gill surfaces into the coelomic fluid. Gills are found in mollusks, annelids, and crustaceans.

Figure 2.4. This common carp, like many other aquatic organisms, has gills that allow it to obtain oxygen from water. (credit: "Guitardude012"/Wikimedia Commons)

The folded surfaces of the gills provide a large surface area to ensure that the fish gets sufficient oxygen. Diffusion is a process in which material travels from regions of high concentration to low concentration until equilibrium is reached. In this case, blood with a low concentration of oxygen molecules circulates through the gills. The concentration of oxygen molecules in water is higher than the concentration of oxygen molecules in gills. As a result, oxygen molecules diffuse from water (high concentration) to blood (low concentration), as shown in <u>Figure 2.5</u>. Similarly, carbon dioxide molecules in the blood diffuse from the blood (high concentration) to water (low concentration).
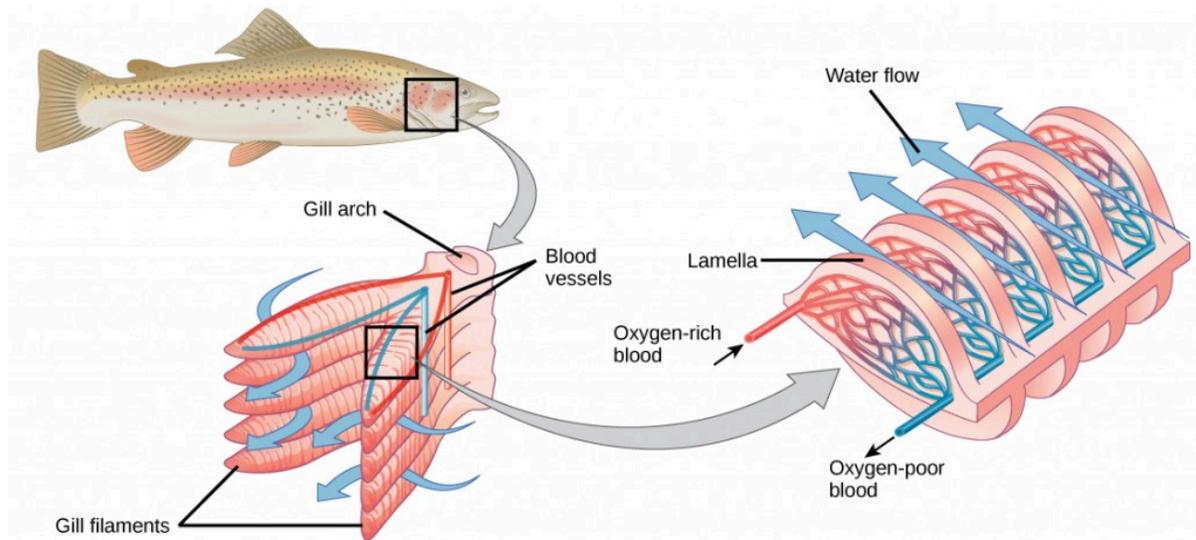
Figure 2.5. As water flows over the gills, oxygen is transferred to blood via the veins. (credit "fish": modification of work by Duane Raver, NOAA)

In terrestrial environments, animals showcase a diversity of respiratory system adaptations to obtaining oxygen (Figure 2.6), such as treachea in insects, alveoli in mammals and parabronchi in birds. You can also use the following gif animation as an overview of how air moves through respiratory system structured of a few terrestrial animals.
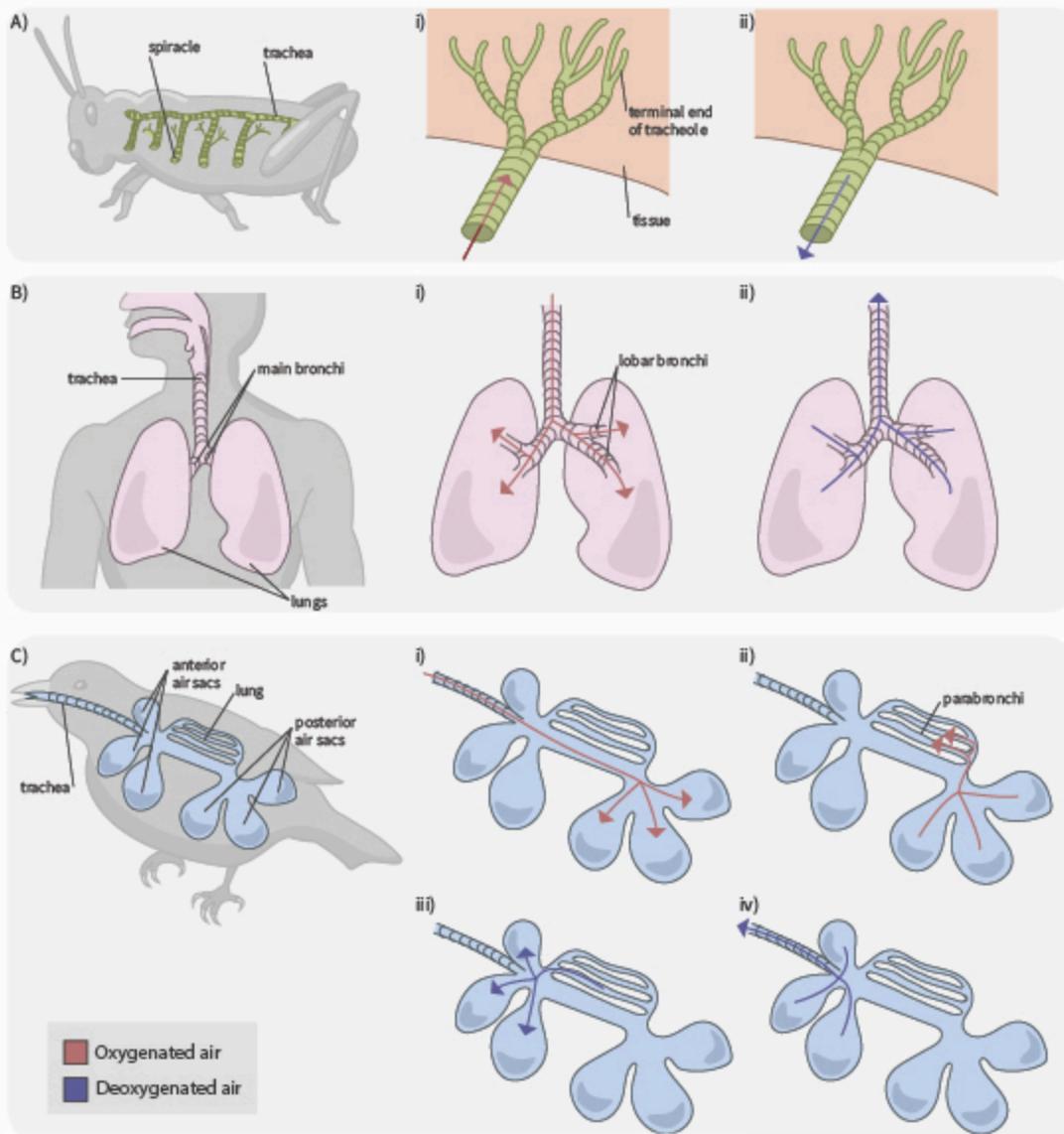
A) *Insects*: i) Oxygenated air enters through open spiracles into trachea, which branch into smaller tubes called tracheoles. The tracheoles directly contact tissue.
ii) Deoxygenated air exits through the tracheoles, to the trachea and out through open spiracles.

B) *Mammals*: i) Inhalation: Oxygenated air enters through the trachea, to the primary bronchi, into the lungs.
ii) Exhalation: Deoxygenated air exits the lungs through the primary bronchi, and out through the trachea.

C) *Birds*: i) First inhalation: Oxygenated air enters through the trachea into the posterior air sacs.
ii) First exhalation: Oxygenated air from the posterior air sacs moves into the lungs through parabronchi.
iii) Second inhalation: Deoxygenated air from the lungs moves to the anterior air sacs.
iv) Second exhalation: Deoxygenated air from the anterior air sacs exits through the trachea.

Figure 2.6. Breathing cycle in insects, humans, and birds. This figure was created and designed by Taeah Kim, a graduate student in Biomedical Communications Masters program at the University of Toronto Mississauga.

Tracheal systems

Insect respiration is independent of its circulatory system; therefore, the blood does not play a direct role in oxygen transport. Insects have a highly specialized type of respiratory system called the tracheal system, which consists of a network of small tubes that carries oxygen to the entire body. The tracheal system is the most direct and efficient respiratory system in active animals. The tubes in the tracheal system are made of a polymeric material called chitin.

Insect bodies have openings, called spiracles along the thorax and abdomen. These openings connect to the tubular network, allowing oxygen to pass into the body ([Figure 2.7](#)) and regulating the diffusion of CO_2 and water vapor. Air enters and leaves the tracheal system through the spiracles. Some insects can ventilate the tracheal system with body movements.
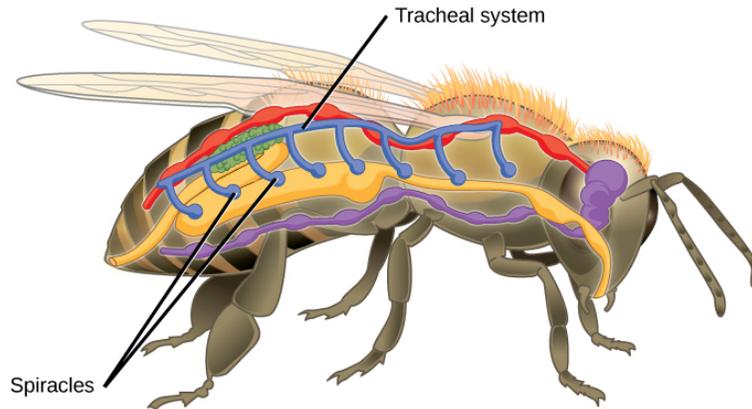
Figure 2.7. Insects perform respiration via a tracheal system.

Mammalian systems

In mammals, pulmonary ventilation occurs via inhalation (breathing). During inhalation, air enters the body through the nasal **cavity** located just inside the nose ([Figure 2.8](#)). As air passes through the nasal cavity, the air is warmed to body temperature and humidified. The respiratory tract is coated with mucus to seal the tissues from direct contact with air. Mucus is high in water. As air crosses these surfaces of the mucous membranes, it picks up water. These processes help equilibrate the air to the body conditions, reducing any damage that cold, dry air can cause. Particulate matter that is floating in the air is removed in the nasal passages via mucus and cilia. The processes of warming, humidifying, and removing particles are important protective mechanisms that prevent damage to the trachea and lungs. Thus, inhalation serves several purposes in addition to bringing oxygen into the respiratory system.
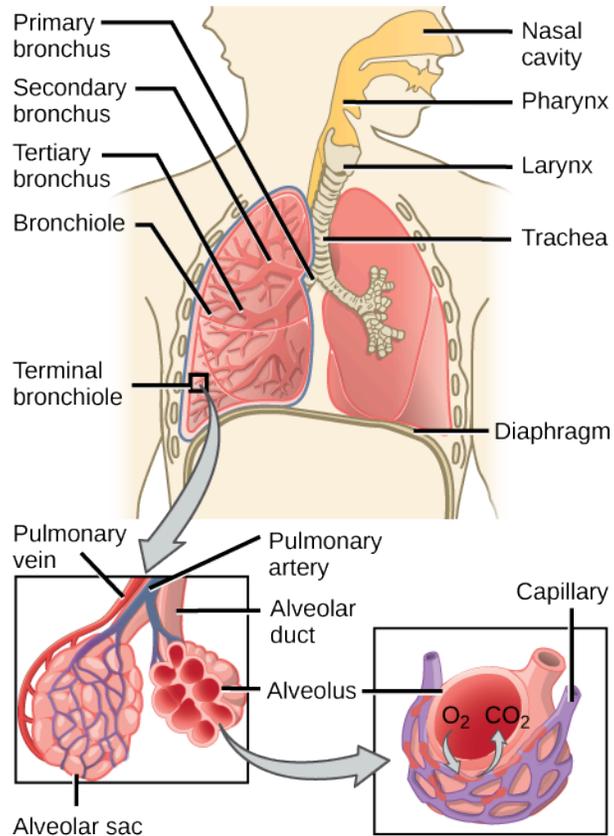
Figure 2.8. Air enters the respiratory system through the nasal cavity and pharynx and then passes through the trachea and into the bronchi, which bring air into the lungs. (credit: modification of work by NCI)

Question 2.4

Which of the following statements about the mammalian respiratory system is false?
a. When we breathe in, air travels from the pharynx to the trachea.
b. The bronchioles branch into bronchi.
c. Alveolar ducts connect to alveolar sacs.
d. Gas exchange between the lung and blood takes place in the alveolus

From the nasal cavity, air passes through the **pharynx** (throat) and the **larynx** (voice box), as it makes its way to the trachea (Figure 2.9). The main function of the trachea is to funnel the inhaled air to the lungs and the exhaled air back out of the body. The human trachea is a cylinder about 10 to 12 cm

long and 2 cm in diameter that sits in front of the esophagus and extends from the larynx into the chest cavity where it divides into the two primary bronchi at the mid-thorax. It is made of incomplete rings of hyaline cartilage and smooth muscle (Figure 2.9). The trachea is lined with mucus-producing goblet cells and ciliated epithelia. The cilia propel foreign particles trapped in the mucus toward the pharynx. The cartilage provides strength and support to the trachea to keep the passage open. The smooth muscle can contract, decreasing the trachea's diameter, which causes expired air to rush upwards from the lungs at a great force. The forced exhalation helps to expel mucus when we cough. Smooth muscle can contract or relax, depending on stimuli from the external environment or the body's nervous system.
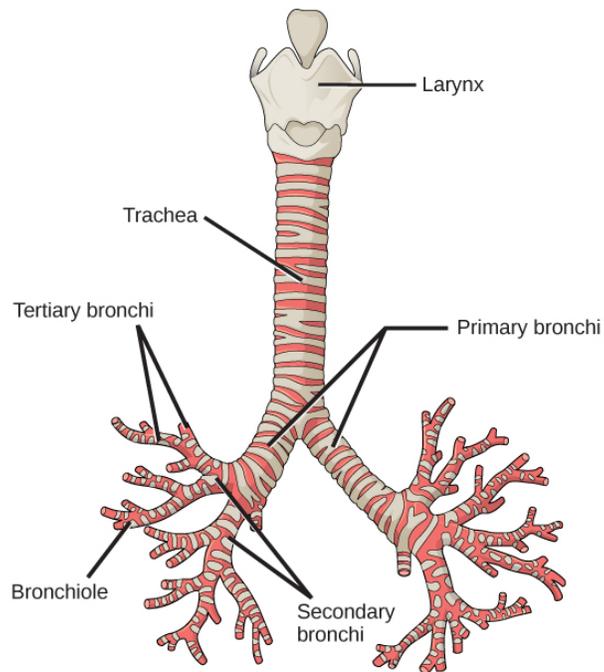
Figure 2.9. The trachea and bronchi are made of incomplete rings of cartilage. (credit: modification of work by Gray's Anatomy)

Lungs: bronchi and alveoli

The end of the trachea bifurcates (divides) to the right and left lungs. The lungs are not identical. The right lung is larger and contains three lobes, whereas the smaller left lung contains two lobes (Figure 2.10). The muscular **diaphragm,** which facilitates breathing, is inferior (below) to the lungs and marks the end of the thoracic cavity.
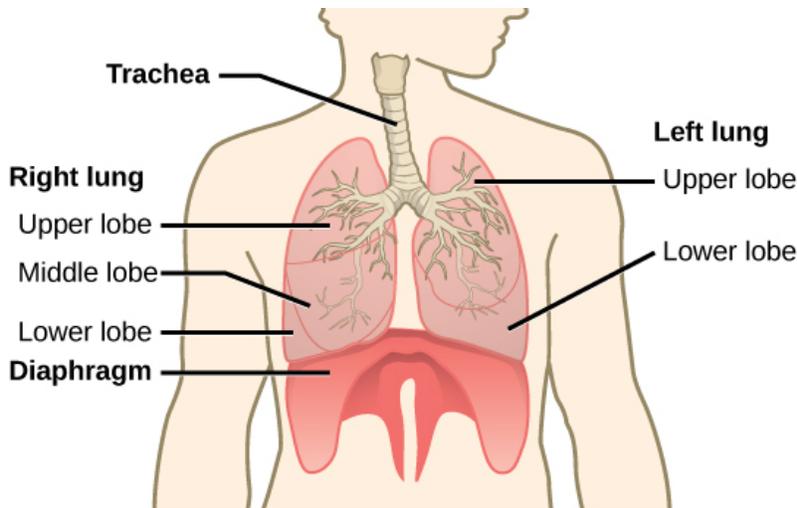
Figure 2.10. The trachea bifurcates into the right and left bronchi in the lungs. The right lung is made of three lobes and is larger. To accommodate the heart, the left lung is smaller and has only two lobes.

In the lungs, air is diverted into smaller and smaller passages or **bronchi.** Air enters the lungs through the two **primary (main) bronchi** (singular: bronchus). Each bronchus divides into secondary bronchi, then into tertiary bronchi, which in turn divide, creating smaller and smaller diameter **bronchioles** as they split and spread through the lung. Like the trachea, the bronchi are made of cartilage and smooth muscle. At the bronchioles, the cartilage is replaced with elastic fibers. Bronchi are innervated by nerves of both the parasympathetic and sympathetic nervous systems that control muscle contraction (parasympathetic) or relaxation (sympathetic) in the bronchi and bronchioles, depending on the nervous system's cues. In humans, bronchioles with a diameter smaller than 0.5 mm are the **respiratory bronchioles**. They lack cartilage and therefore rely on inhaled air to support their shape. As the passageways decrease in diameter, the relative amount of smooth muscle increases.

The **terminal bronchioles** subdivide into microscopic branches called respiratory bronchioles. The respiratory bronchioles subdivide into several alveolar ducts. Numerous alveoli and alveolar sacs surround the alveolar ducts. The alveolar sacs resemble bunches of grapes tethered to the end of the bronchioles (Figure 2.11). In the acinar region, the **alveolar ducts** are attached to the end of each bronchiole. At the end of each duct are approximately 100 **alveolar sacs,** each containing 20 to 30 **alveoli** that are 200 to 300 microns in diameter. Gas exchange occurs only in alveoli. Alveoli are made of thin-walled parenchymal cells, typically one-cell thick, that look like tiny bubbles within the sacs. Alveoli are in direct contact with capillaries (one-cell thick) of the circulatory system. Such intimate contact ensures that oxygen will diffuse from alveoli into the blood and be distributed to the cells of the body. In addition, the carbon dioxide that was produced by cells as a waste product will diffuse from the blood into alveoli to be exhaled. The anatomical arrangement of capillaries and alveoli emphasizes the structural and functional relationship of the respiratory and circulatory systems. Because there are so many alveoli (~300 million per lung) within each alveolar sac and so many sacs at

the end of each alveolar duct, the lungs have a sponge-like consistency. This organization produces a very large surface area that is available for gas exchange. The surface area of alveoli in the lungs is approximately 75 m2. This large surface area, combined with the thin-walled nature of the alveolar parenchymal cells, allows gases to easily diffuse across the cells.
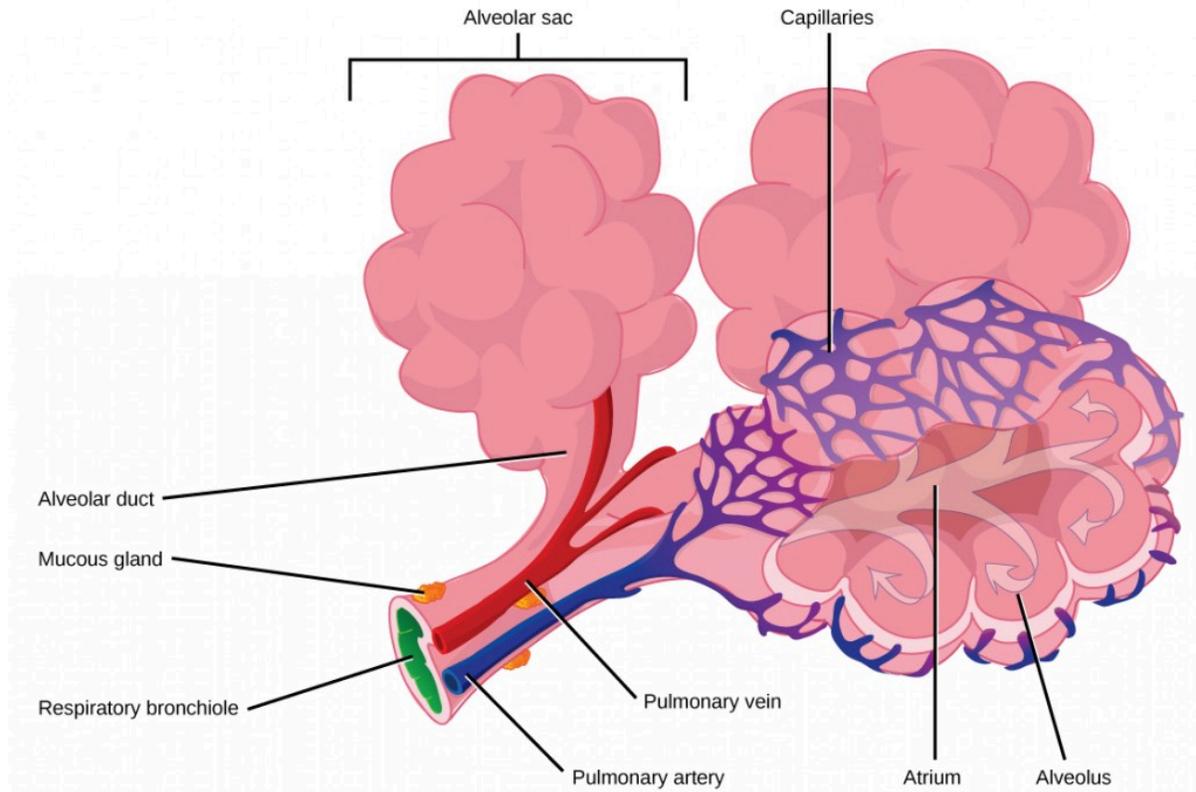
Figure 2.11. Terminal bronchioles are connected by respiratory bronchioles to alveolar ducts and alveolar sacs. Each alveolar sac contains 20 to 30 spherical alveoli and has the appearance of a bunch of grapes. Air flows into the atrium of the alveolar sac, then circulates into alveoli where gas exchange occurs with the capillaries. Mucous glands secrete mucous into the airways, keeping them moist and flexible. (credit: modification of work by Mariana Ruiz Villareal)

Question 2.5

Which of the following statements about the human respiratory system is false?
a. When we breathe in, air travels from the pharynx to the trachea.
b. The bronchioles branch into bronchi.
c. Alveolar ducts connect to alveolar sacs.
d. Gas exchange between the lungs and blood takes place in the alveolus.

For further interest and comparison of human lungs with other animal respiratory systems, here is a link to an [interesting article](#).

Pulmonary ventilation is the act of breathing, which can be described as the movement of air into and out of the lungs. The major mechanisms that drive pulmonary ventilation are atmospheric pressure (P_{atm}); the air pressure within the alveoli, called intra-alveolar pressure (P_{alv}); and the pressure within the pleural cavity, called intrapleural pressure (P_{ip}).

Mechanisms of breathing

The intra-alveolar and intrapleural pressures are dependent on certain physical features of the lung. However, the ability to breathe—to have air enter the lungs during inspiration and air leave the lungs during expiration—is dependent on the air pressure of the atmosphere and the air pressure within the lungs.

Pressure relationships

Inspiration (or inhalation) and expiration (or exhalation) are dependent on the differences in pressure between the atmosphere and the lungs. In a gas, the pressure is a force created by the movement of gas molecules that are confined. For example, a certain number of gas molecules in a two-liter container has more room than the same number of gas molecules in a one-liter container ([Figure 2.13](#)). In this case, the force exerted by the movement of the gas molecules against the walls of the two-liter container is lower than the force exerted by the gas molecules in the one-liter container. Therefore, the pressure is lower in the two-liter container and higher in the one-liter container. At a constant temperature, changing the volume occupied by the gas changes the pressure, as does changing the number of gas molecules. Boyle's law describes the relationship between volume and pressure in a gas at a constant temperature. Boyle discovered that the pressure of a gas is inversely proportional to its volume: If volume increases, pressure decreases. Likewise, if volume decreases, pressure increases. Pressure and volume are inversely related ($P = k/V$). Therefore, the pressure in the one-liter container (one-half the volume of the two-liter container) would be twice the pressure in the two-liter container. Boyle's law is expressed by the following formula:

$$P_1V_1 = P_2V_2 P_1V_1 = P_2V_2$$

In this formula, P_1 represents the initial pressure and V_1 represents the initial volume, whereas the final pressure and volume are represented by P_2 and V_2, respectively. If the two- and one-liter containers were connected by a tube and the volume of one of the containers were changed, then the gases would move from higher pressure (lower volume) to lower pressure (higher volume).

As volume increases, pressure decreases

As volume decreases, pressure increases

$$pV = k$$
$$p_1V_1 = p_2V_2$$
$$p = k/V$$
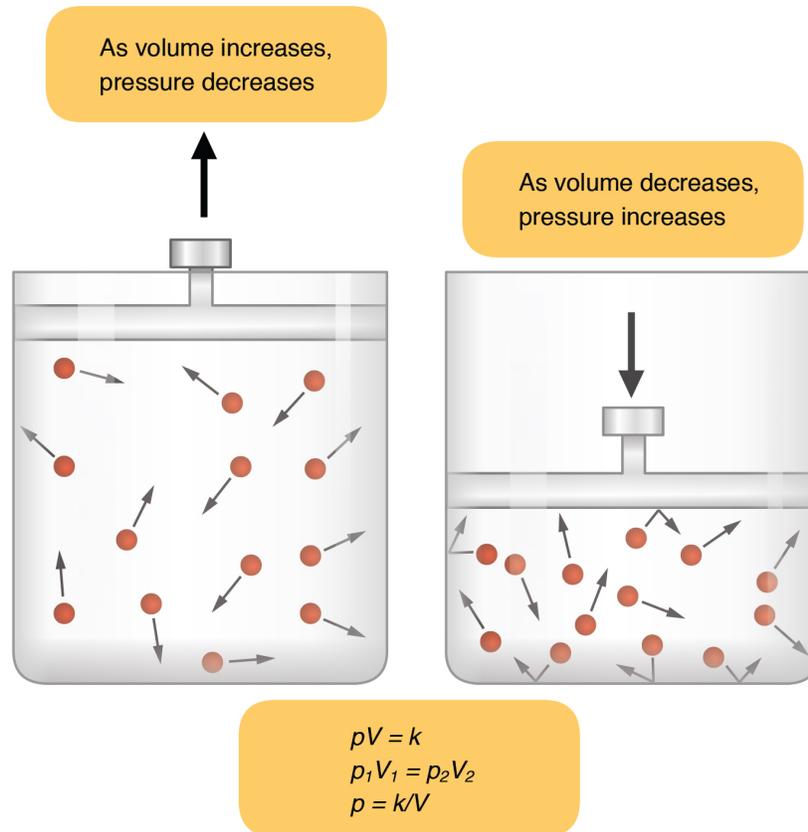
Figure 2.13. Boyle's law. In a gas, pressure increases as volume decreases.

Pulmonary ventilation is dependent on three types of pressure: atmospheric, intra-alveolar, and intrapleural. Atmospheric pressure is the amount of force that is exerted by gases in the air surrounding any given surface, such as the body. Atmospheric pressure can be expressed in terms of the unit atmosphere, abbreviated atm, or in millimeters of mercury (mm Hg). One atm is equal to 760 mm Hg, which is the atmospheric pressure at sea level. Typically, for respiration, other pressure values are discussed in relation to atmospheric pressure. Therefore, negative pressure is pressure lower than the atmospheric pressure, whereas positive pressure is the pressure that it is greater than the atmospheric pressure. A pressure that is equal to the atmospheric pressure is expressed as zero.

Intra-alveolar pressure (intrapulmonary pressure) is the pressure of the air within the alveoli, which changes during the different phases of breathing (Figure 2.14). Because the alveoli are connected to the atmosphere via the tubing of the airways (similar to the two- and one-liter containers in the

example above), the intrapulmonary pressure of the alveoli always equalizes with the atmospheric pressure.

Intrapulmonary and Intrapleural Pressure Relationships

Atmospheric pressure

Parietal pleura

Visceral pleura

Transpulmonary pressure:
760 mm Hg –756 mm Hg
= 4 mm Hg

Pleural cavity

Intrapleural pressure:
756 mm Hg (–4 mm Hg)

Thoracic wall

Lung

Intra-alveolar pressure:
760 mm Hg (0 mm Hg)

Diaphragm
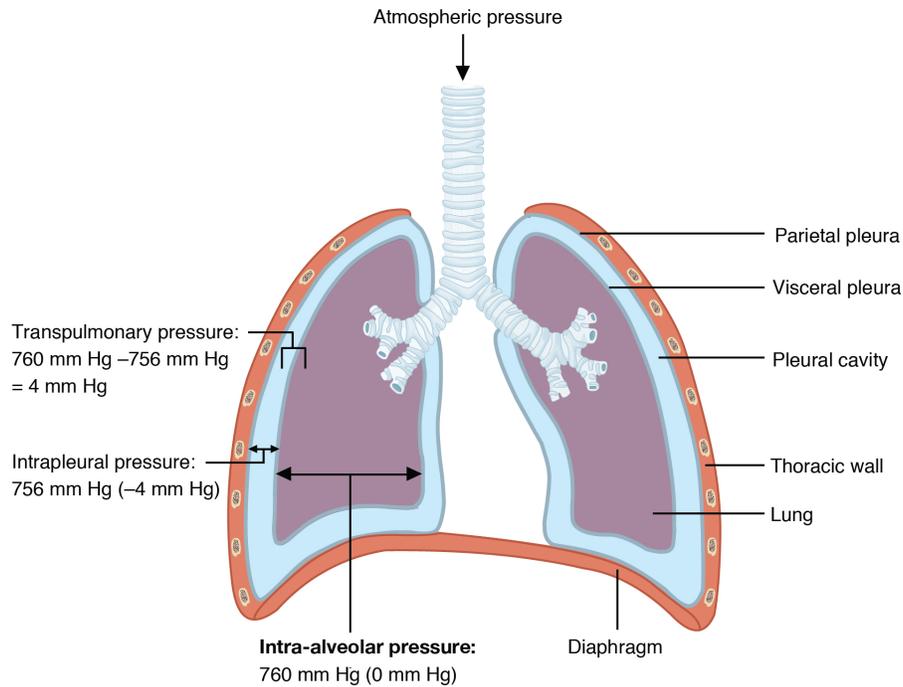
Figure 2.14. Intra-alveolar pressure changes during the different phases of the cycle. It equalizes at 760 mm Hg but does not remain at 760 mm Hg.

Intrapleural pressure is the pressure of the air within the pleural cavity, between the visceral and parietal pleurae. Similar to intra-alveolar pressure, intrapleural pressure also changes during the different phases of breathing. However, due to certain characteristics of the lungs, the intrapleural pressure is always lower than, or negative to, the intra-alveolar pressure (and therefore also to atmospheric pressure). Although it fluctuates during inspiration and expiration, intrapleural pressure remains approximately –4 mm Hg throughout the breathing cycle.

Competing forces within the thorax cause the formation of the negative intrapleural pressure. One of these forces relates to the elasticity of the lungs themselves–elastic tissue pulls the lungs inward, away from the thoracic wall. Surface tension of the alveolar fluid, which is mostly water, also creates an inward pull of the lung tissue. This inward tension from the lungs is countered by opposing forces from the pleural fluid and thoracic wall. Surface tension within the pleural cavity pulls the lungs outward. Too much or too little pleural fluid would hinder the creation of the negative intrapleural pressure; therefore, the level must be closely monitored by the mesothelial cells and drained by the lymphatic system. Since the parietal pleura is attached to the thoracic wall, the natural elasticity of the chest wall opposes the inward pull of the lungs. Ultimately, the outward pull is slightly greater than the inward pull, creating the –4 mm Hg intrapleural pressure relative to the intra-alveolar

pressure. Transpulmonary pressure is the difference between the intrapleural and intra-alveolar pressures, and it determines the size of the lungs. A higher transpulmonary pressure corresponds to a larger lung.

Physical factors affecting ventilation

In addition to the differences in pressures, breathing is also dependent upon the contraction and relaxation of muscle fibers of both the diaphragm and thorax. The lungs themselves are passive during breathing, meaning they are not involved in creating the movement that helps inspiration and expiration. This is because of the adhesive nature of the pleural fluid, which allows the lungs to be pulled outward when the thoracic wall moves during inspiration. The recoil of the thoracic wall during expiration causes compression of the lungs. Contraction and relaxation of the diaphragm and intercostals muscles (found between the ribs) cause most of the pressure changes that result in inspiration and expiration. These muscle movements and subsequent pressure changes cause air to either rush in or be forced out of the lungs.

Other characteristics of the lungs influence the effort that must be expended to ventilate. Resistance is a force that slows motion, in this case, the flow of gases. The size of the airway is the primary factor affecting resistance. A small tubular diameter forces air through a smaller space, causing more collisions of air molecules with the walls of the airways. The following formula helps to describe the relationship between airway resistance and pressure changes:

$$F = \Delta P / R F = \Delta P / R$$

As noted earlier, there is surface tension within the alveoli caused by water present in the lining of the alveoli. This surface tension tends to inhibit expansion of the alveoli. However, pulmonary surfactant secreted by type II alveolar cells mixes with that water and helps reduce this surface tension. Without pulmonary surfactant, the alveoli would collapse during expiration.

Thoracic wall compliance is the ability of the thoracic wall to stretch while under pressure. This can also affect the effort expended in the process of breathing. In order for inspiration to occur, the thoracic cavity must expand. The expansion of the thoracic cavity directly influences the capacity of the lungs to expand. If the tissues of the thoracic wall are not very compliant, it will be difficult to expand the thorax to increase the size of the lungs.

Gas laws and air composition

Gas molecules exert a force on the surfaces with which they are in contact; this force is called pressure. In natural systems, gases are normally present as a mixture of different types of molecules.

For example, the atmosphere consists of oxygen, nitrogen, carbon dioxide, and other gaseous molecules, and this gaseous mixture exerts a certain pressure referred to as atmospheric pressure (Table 2.1). Partial pressure (P_x) is the pressure of a single type of gas in a mixture of gases. For example, in the atmosphere, oxygen exerts a partial pressure, and nitrogen exerts another partial pressure, independent of the partial pressure of oxygen (Figure 2.15). Total pressure is the sum of all the partial pressures of a gaseous mixture. Dalton's law describes the behavior of nonreactive gases in a gaseous mixture and states that a specific gas type in a mixture exerts its own pressure; thus, the total pressure exerted by a mixture of gases is the sum of the partial pressures of the gases in the mixture.

Table 1. Partial and total pressures of a gas.

Partial Pressures of Atmospheric Gases

Gas	Percent of total composition	Partial pressure (mm Hg)
Nitrogen (N_2)	78.6	597.4
Oxygen (O_2)	20.9	158.8
Water (H_2O)	0.4	3.0
Carbon dioxide (CO_2)	0.04	0.3
Others	0.06	0.5
Total composition/total atmospheric pressure	100%	760.0
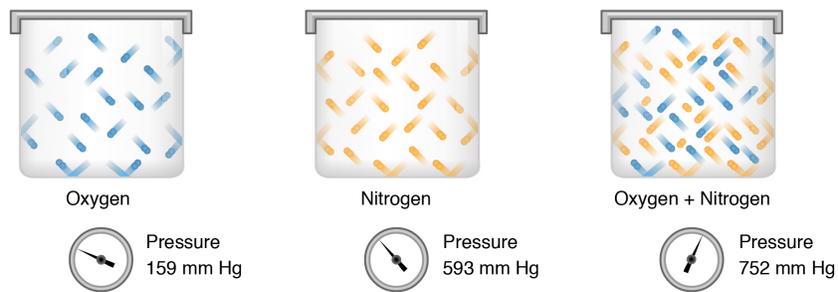
Figure 2.15. Partial pressure is the force exerted by a gas. The sum of the partial pressures of all the gases in a mixture equals the total pressure.

Partial pressure is extremely important in predicting the movement of gases. Recall that gases tend to equalize their pressure in two regions that are connected. A gas will move from an area where its partial pressure is higher to an area where its partial pressure is lower. In addition, the greater the partial pressure difference between the two areas, the more rapid is the movement of gases.

Solubility of gases in liquids

Henry's law describes the behavior of gases when they come into contact with a liquid, such as blood. Henry's law states that the concentration of gas in a liquid is directly proportional to the solubility and partial pressure of that gas. The greater the partial pressure of the gas, the greater the number of gas molecules that will dissolve in the liquid. The concentration of the gas in a liquid is also dependent on the solubility of the gas in the liquid. For example, although nitrogen is present in the atmosphere, very little nitrogen dissolves into the blood because the solubility of nitrogen in blood is very low. The exception to this occurs in scuba divers; the composition of the compressed air that divers breathe causes nitrogen to have a higher partial pressure than normal, causing it to dissolve in the blood in greater amounts than normal. Too much nitrogen in the bloodstream results in a serious condition that can be fatal if not corrected. Gas molecules establish an equilibrium between those molecules dissolved in liquid and those in the air.

The composition of air in the atmosphere and in the alveoli differs. In both cases, the relative concentration of gases is nitrogen > oxygen > water vapor > carbon dioxide. The amount of water vapor present in the alveolar air is greater than that in atmospheric air (Table 2.2). Recall that the respiratory system works to humidify incoming air, thereby causing the air present in the alveoli to have a greater amount of water vapor than atmospheric air. In addition, alveolar air contains a greater amount of carbon dioxide and less oxygen than atmospheric air. This is no surprise, as gas exchange removes oxygen from and adds carbon dioxide to the alveolar air. Both deep and forced breathing cause the alveolar air composition to be changed more rapidly than during quiet breathing. As a result, the partial pressures of oxygen and carbon dioxide change, affecting the diffusion process that moves these materials across the membrane. This will cause oxygen to enter and carbon dioxide to leave the blood more quickly.

Table 2.2. Partial and total pressures of alveolar gases.

Composition and Partial Pressures of Alveolar Air

Gas	Percent of total composition	Partial pressure (mm Hg)
Nitrogen (N_2)	74.9	569
Oxygen (O_2)	13.7	104
Water (H_2O)	6.2	40
Carbon dioxide (CO_2)	5.2	47
Total composition/total alveolar pressure	100%	760.0

Overall, the partial pressure of oxygen in alveolar air is about 104 mm Hg, whereas the partial pressure of the oxygenated pulmonary venous blood is about 100 mm Hg. When ventilation is sufficient, oxygen enters the alveoli at a high rate, and the partial pressure of oxygen in the alveoli remains

high. In contrast, when ventilation is insufficient, the partial pressure of oxygen in the alveoli drops. Without the large difference in partial pressure between the alveoli and the blood, oxygen does not diffuse efficiently across the respiratory membrane. The body has mechanisms that counteract this problem. In cases when ventilation is not sufficient for an alveolus, the body redirects blood flow to alveoli that are receiving sufficient ventilation. This is achieved by constricting the pulmonary arterioles that serves the dysfunctional alveolus, which redirects blood to other alveoli that have sufficient ventilation. At the same time, the pulmonary arterioles that serve alveoli receiving sufficient ventilation vasodilate, which brings in greater blood flow. Factors such as carbon dioxide, oxygen, and pH levels can all serve as stimuli for adjusting blood flow in the capillary networks associated with the alveoli.

Ventilation is regulated by the diameter of the airways, whereas perfusion is regulated by the diameter of the blood vessels. The diameter of the bronchioles is sensitive to the partial pressure of carbon dioxide in the alveoli. A greater partial pressure of carbon dioxide in the alveoli causes the bronchioles to increase their diameter as will a decreased level of oxygen in the blood supply, allowing carbon dioxide to be exhaled from the body at a greater rate. As mentioned above, a greater partial pressure of oxygen in the alveoli causes the pulmonary arterioles to dilate, increasing blood flow.

Question 2.6

Gas moves from an area of _____ partial pressure to an area of _____ partial pressure.
a. low; high
b. low; low
c. high; high
d. high; low

Question 2.7

The partial pressure of carbon dioxide is 45 mm Hg in the blood and 40 mm Hg in the alveoli. What happens to the carbon dioxide?
a. It diffuses into the blood.
b. It diffuses into the alveoli.
c. The gradient is too small for carbon dioxide to diffuse.
d. It decomposes into carbon and oxygen.

Question 2.8
A smoker develops damage to several alveoli that then can no longer function. How does this affect gas exchange?

2.3 – Gaseous Exchange Mechanisms

2.4. Compare counter- and crosscurrent exchange using specific respiratory system examples.

Most gas exchange in the respiratory system structures in animals takes place in counter-current arrangement. In short, this means that the medium that delivers the oxygen (air or water) and the structures that the oxygen is delivered to are moving in directions opposite to one another. An example of this is a fish gill system (Figure 2.12). Another way that gas exchange can happen is through cross-current exchange in which the air carrying the oxygen is moving through the respiratory structures (parabronchi), which are positined perpendicular to the movement of blood around the parabronchi. ([Figure 2.12](#)). Please note that in the figure below, the oxygen-rich and oxygen-poor blood flow direction with respect to the secondary lamellae should be reversed (Figure 2.12a). Additonally, the flow of blood within the barabronchus (Figure 2.12b, direction flow should be from oxygen-poor to oxygen-rich blood). The open textbook developers and author are currently working on updating this figure.
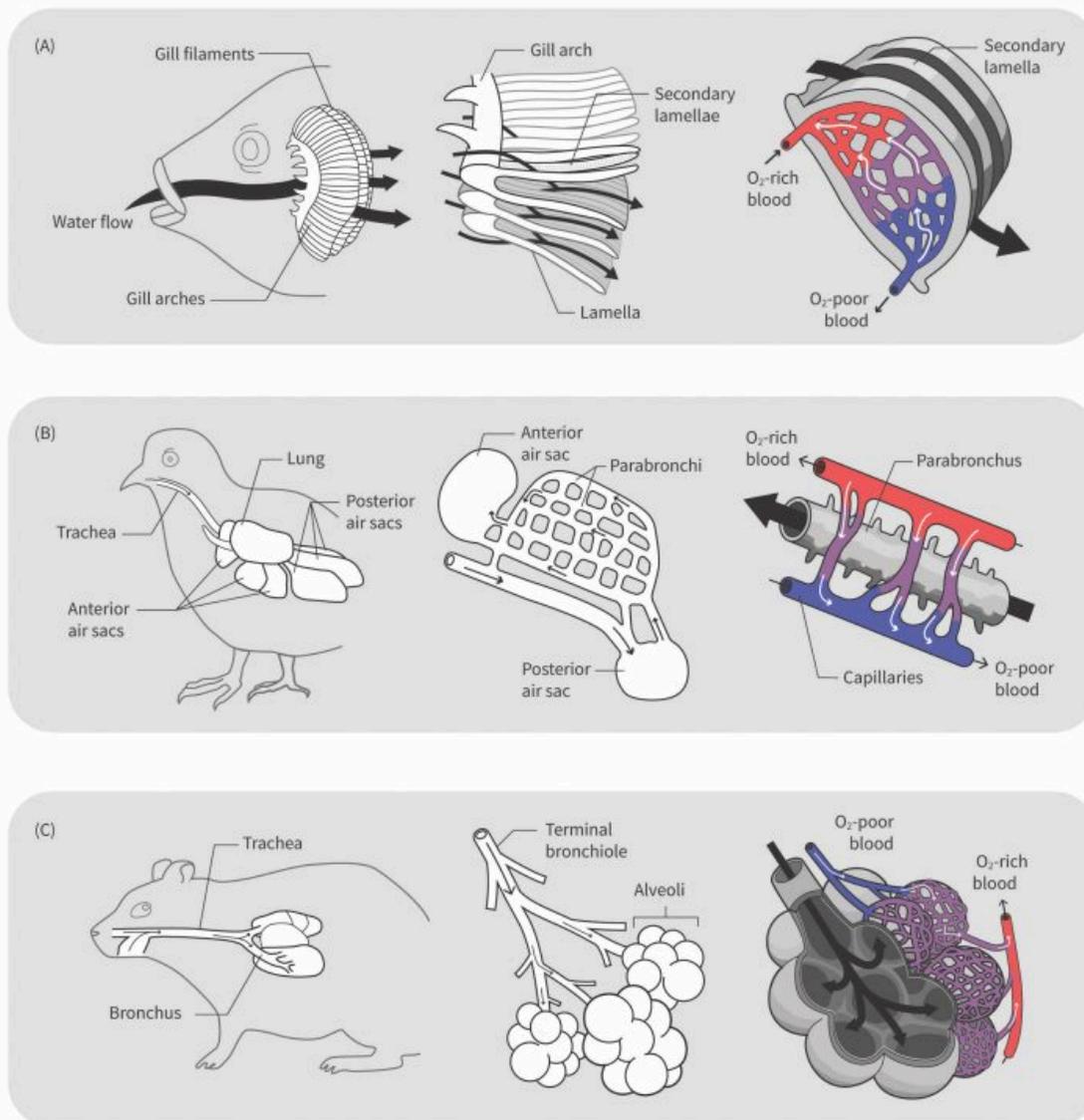
. Comparison of gas exchange in fish gills, bird parabronchi, and mammalian lung (A) The gills of teleost fish consists of gill arches that support gill filaments. Gill filaments have a number of gill lamellae that deliver water to the capillary network in the gill arches. Blood flow through the capillaries of the secondary lamella is arranged in a counter-current pattern relative to the flow of the water through the gills. (B) In birds, a series of flexible air sacs work together with the lung. Within the lung, there are parallel small tubes called the parabronchi. Blood flow through the air capillaries is arranged in a cross-current pattern relative to the air flow across the parabronchi. (C) Air enters the mammalian lung via the mouth and nares, and then enters the trachea. The trachea branches into primary bronchi, which branches into successively smaller tubes. The bronchioles, the smallest tubes, terminate in sacs called alveoli where gas exchange occurs.

Figure 2.12. Comparison of gas exchange in gills, parabronchi, and alveoli. This figure was created and designed by Christine Shan, a graduate student in Biomedical Communications Masters program at the University of Toronto Mississauga (developed as a project in MSC2001Y Visual Representation of Medical Knowledge course in the BMC

MSc program).

Question 2.9.

Which one do you think is an example of counter-current and which one is an example of crosscurrent exchange?

CHAPTER 3 – TRANSPORT OF GASES

3.1. Explain how oxygen and carbon dioxide are transported in blood or equivalent fluid medium; and

3.2. Compare and contrast different circulatory systems using specific animal examples and evolution.

3.1 – Gas Transport

The content of this chapter was adapted from the *Anatomy and Physiology* open textbook ([Chapter 22.4 – Gas Exchange](#)).

3.1. Explain how oxygen and carbon dioxide are transported in blood or fluid medium.

The other major activity in the lungs is the process of respiration, the process of gas exchange. The function of respiration is to provide oxygen for use by body cells during cellular respiration and to eliminate carbon dioxide, a waste product of cellular respiration, from the body. In order for the exchange of oxygen and carbon dioxide to occur, both gases must be transported between the external and internal respiration sites. Although carbon dioxide is more soluble than oxygen in blood, both gases require a specialized transport system for the majority of the gas molecules to be moved between the lungs and other tissues.

Therefore, gas exchange occurs at two sites in the body: in the lungs, where oxygen is picked up and carbon dioxide is released at the respiratory membrane, and at the tissues, where oxygen is released and carbon dioxide is picked up. External respiration is the exchange of gases with the external environment and occurs in the alveoli of the lungs. Internal respiration is the exchange of gases with the internal environment and occurs in the tissues. The actual exchange of gases occurs due to simple diffusion. Energy is not required to move oxygen or carbon dioxide across membranes. Instead, these gases follow pressure gradients that allow them to diffuse. The anatomy of the lung maximizes the diffusion of gases: The respiratory membrane is highly permeable to gases; the respiratory and blood capillary membranes are very thin, and there is a large surface area throughout the lungs.

External respiration

The pulmonary artery carries deoxygenated blood into the lungs from the heart, where it branches and eventually becomes the capillary network composed of pulmonary capillaries. These pulmonary capillaries create the respiratory membrane with the alveoli ([Figure 3.1](#)). As the blood is pumped

through this capillary network, gas exchange occurs. Although a small amount of the oxygen is able to dissolve directly into the plasma from the alveoli, most of the oxygen is picked up by erythrocytes (red blood cells) and binds to a protein called hemoglobin, a process described later in this chapter. Oxygenated hemoglobin is red, causing the overall appearance of bright red oxygenated blood, which returns to the heart through the pulmonary veins. Carbon dioxide is released in the opposite direction of oxygen, from the blood to the alveoli. Some of the carbon dioxide is returned on hemoglobin, but can also be dissolved in plasma or is present as a converted form, also explained in greater detail later in this chapter.

External respiration occurs as a function of partial pressure differences in oxygen and carbon dioxide between the alveoli and the blood in the pulmonary capillaries.

External Respiration
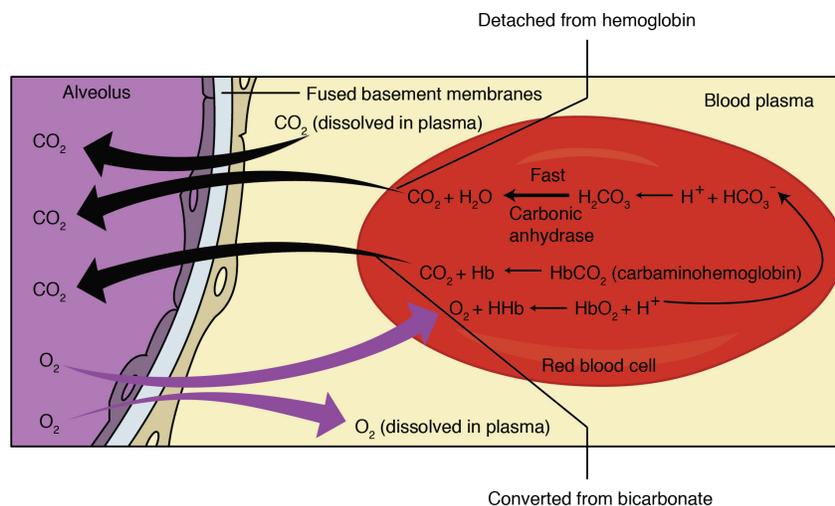
Figure 3.1. In external respiration, oxygen diffuses across the respiratory membrane from the alveolus to the capillary, whereas carbon dioxide diffuses out of the capillary into the alveolus.

Although the solubility of oxygen in the blood is not high, there is a drastic difference in the partial pressure of oxygen in the alveoli versus in the blood of the pulmonary capillaries. This difference is about 64 mm Hg: The partial pressure of oxygen in the alveoli is about 104 mm Hg, whereas its partial pressure in the blood of the capillary is about 40 mm Hg. This large difference in partial pressure creates a very strong pressure gradient that causes oxygen to rapidly cross the respiratory membrane from the alveoli into the blood.

The partial pressure of carbon dioxide is also different between the alveolar air and the blood of the capillary. However, the partial pressure difference is less than that of oxygen, about 5 mm Hg. The partial pressure of carbon dioxide in the blood of the capillary is about 45 mm Hg, whereas its partial pressure in the alveoli is about 40 mm Hg. However, the solubility of carbon dioxide is much greater than that of oxygen—by a factor of about 20—in both blood and alveolar fluids. As a result, the

relative concentrations of oxygen and carbon dioxide that diffuse across the respiratory membrane are similar.

Internal respiration

Internal respiration is the gas exchange that occurs at the level of body tissues ([Figure 3.2](#)). Similar to external respiration, internal respiration also occurs as simple diffusion due to a partial pressure gradient. However, the partial pressure gradients are the opposite of those present at the respiratory membrane. The partial pressure of oxygen in tissues is low, about 40 mm Hg, because oxygen is continuously used for cellular respiration. In contrast, the partial pressure of oxygen in the blood is about 100 mm Hg. This creates a pressure gradient that causes oxygen to dissociate from hemoglobin, diffuse out of the blood, cross the interstitial space, and enter the tissue. Hemoglobin that has little oxygen bound to it loses much of its brightness, so that blood returning to the heart is more burgundy in color.

Considering that cellular respiration continuously produces carbon dioxide, the partial pressure of carbon dioxide is lower in the blood than it is in the tissue, causing carbon dioxide to diffuse out of the tissue, cross the interstitial fluid, and enter the blood. It is then carried back to the lungs either bound to hemoglobin, dissolved in plasma, or in a converted form. By the time blood returns to the heart, the partial pressure of oxygen has returned to about 40 mm Hg, and the partial pressure of carbon dioxide has returned to about 45 mm Hg. The blood is then pumped back to the lungs to be oxygenated once again during external respiration.
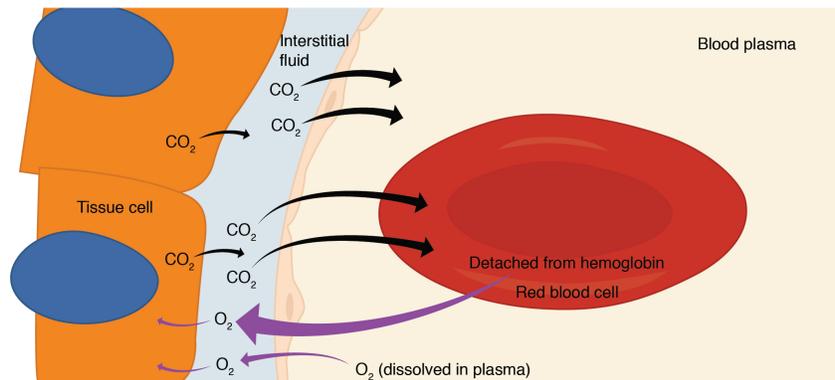
Figure 3.2. Internal respiration: oxygen diffuses out of the capillary and into cells, whereas carbon dioxide diffuses out of cells and into the capillary.

A hyperbaric chamber is a unit that can be sealed and expose a patient to either 100 percent oxygen with increased pressure or a mixture of gases that includes a higher concentration of oxygen than normal atmospheric air, also at a higher partial pressure than the atmosphere ([Figure 3.3](#)).

Hyperbaric chamber treatment is based on the behaviour of gases. As you recall, gases move from a region of higher partial pressure to a region of lower partial pressure. In a hyperbaric chamber, the atmospheric pressure is increased, causing a greater amount of oxygen than normal to diffuse into the bloodstream of the patient. Hyperbaric chamber therapy is used to treat a variety of medical problems, such as wound and graft healing, anaerobic bacterial infections, and carbon monoxide poisoning. Exposure to and poisoning by carbon monoxide is difficult to reverse, because hemoglobin's affinity for carbon monoxide is much stronger than its affinity for oxygen, causing carbon monoxide to replace oxygen in the blood. Hyperbaric chamber therapy can treat carbon monoxide poisoning because the increased atmospheric pressure causes more oxygen to diffuse into the bloodstream. At this increased pressure and increased concentration of oxygen, carbon monoxide is displaced from hemoglobin. Another example is the treatment of anaerobic bacterial infections, which are created by bacteria that cannot or prefer not to live in the presence of oxygen. An increase in blood and tissue levels of oxygen helps to kill the anaerobic bacteria that are responsible for the infection, as oxygen is toxic to anaerobic bacteria. For wounds and grafts, the chamber stimulates the healing process by increasing energy production needed for repair. Increasing oxygen transport allows cells to ramp up cellular respiration and thus ATP production, the energy needed to build new structures.

Figure 3.3. Hyperbaric Chamber (credit: "komunews"/flickr.com)

Oxygen transport in the blood

Even though oxygen is transported via the blood, you may recall that oxygen is not very soluble in liquids. A small amount of oxygen does dissolve in the blood and is transported in the bloodstream, but it is only about 1.5% of the total amount. The majority of oxygen molecules are carried from the

lungs to the body's tissues by a specialized transport system, which relies on the erythrocyte–the red blood cell. Erythrocytes contain a metalloprotein, hemoglobin, which serves to bind oxygen molecules to the erythrocyte ([Figure 3.4](#)). Heme is the portion of hemoglobin that contains iron, and it is the heme that binds the oxygen. One hemoglobin molecule contains four iron-containing hemes, and because of this, each hemoglobin molecule is capable of carrying up to four molecules of oxygen. As oxygen diffuses across the respiratory membrane from the alveolus to the capillary, it also diffuses into the red blood cell and is bound by hemoglobin. The following reversible chemical reaction describes the production of the final product, oxyhemoglobin ($Hb–O_2$), which is formed when oxygen binds to hemoglobin. Oxyhemoglobin is a bright red-colored molecule that contributes to the bright red color of oxygenated blood.

$Hb + O2 \leftrightarrow Hb – O2Hb + O2 \leftrightarrow Hb – O2$

In this formula, Hb represents reduced hemoglobin, that is, hemoglobin that does not have oxygen bound to it. There are multiple factors involved in how readily heme binds to and dissociates from oxygen, which will be discussed in the subsequent sections.

Erythrocyte and Hemoglobin
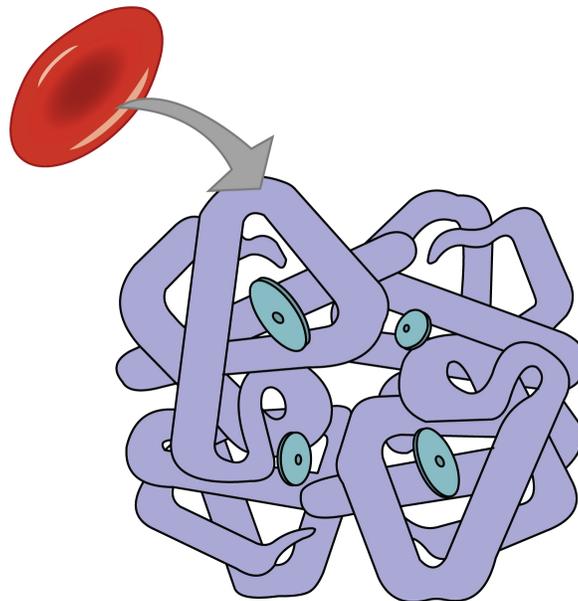
*Figure 3.4. Hemoglobin consists of four subunits, each of which contains one molecule of iron.*

Function of hemoglobin

Hemoglobin is composed of subunits, a protein structure that is referred to as a quaternary structure. Each of the four subunits that make up hemoglobin is arranged in a ring-like fashion, with an iron atom covalently bound to the heme in the center of each subunit. Binding of the first oxygen molecule

causes a conformational change in hemoglobin that allows the second molecule of oxygen to bind more readily. As each molecule of oxygen is bound, it further facilitates the binding of the next molecule, until all four heme sites are occupied by oxygen. The opposite occurs as well: After the first oxygen molecule dissociates and is "dropped off" at the tissues, the next oxygen molecule dissociates more readily. When all four heme sites are occupied, the hemoglobin is said to be saturated. When one to three heme sites are occupied, the hemoglobin is said to be partially saturated. Therefore, when considering the blood as a whole, the percent of the available heme units that are bound to oxygen at a given time is called hemoglobin saturation. Hemoglobin saturation of 100 percent means that every heme unit in all of the erythrocytes of the body is bound to oxygen. In a healthy individual with normal hemoglobin levels, hemoglobin saturation generally ranges from 95 percent to 99 percent.

Oxygen dissociation from hemoglobin

Partial pressure is an important aspect of the binding of oxygen to and disassociation from heme. An oxygen-hemoglobin dissociation curve is a graph that describes the relationship of partial pressure to the binding of oxygen to heme and its subsequent dissociation from heme (Figure 3.5). Remember that gases travel from an area of higher partial pressure to an area of lower partial pressure. In addition, the affinity of an oxygen molecule for heme increases as more oxygen molecules are bound. Therefore, in the oxygen-hemoglobin saturation curve, as the partial pressure of oxygen increases, a proportionately greater number of oxygen molecules are bound by heme. Not surprisingly, the oxygen-hemoglobin saturation/dissociation curve also shows that the lower the partial pressure of oxygen, the fewer oxygen molecules are bound to heme. As a result, the partial pressure of oxygen plays a major role in determining the degree of binding of oxygen to heme at the site of the respiratory membrane, as well as the degree of dissociation of oxygen from heme at the site of body tissues.
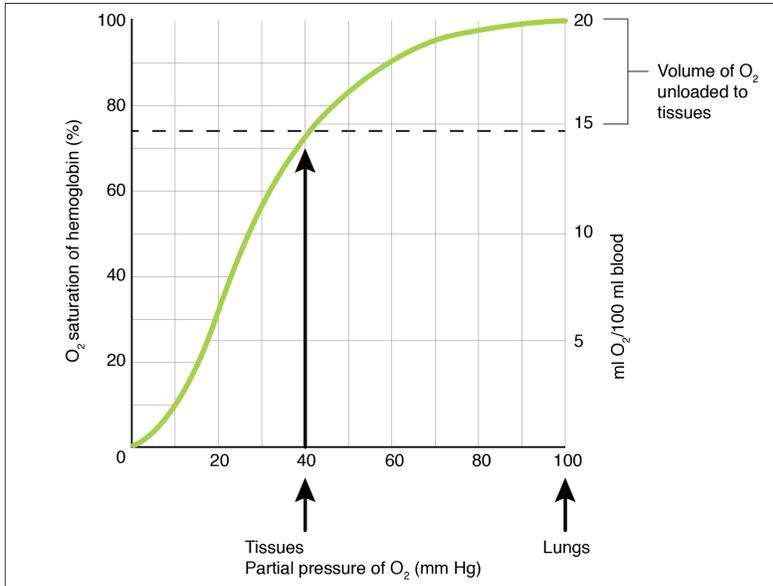
(a) Partial pressure of oxygen and hemoglobin saturation

(b) Effect of pH
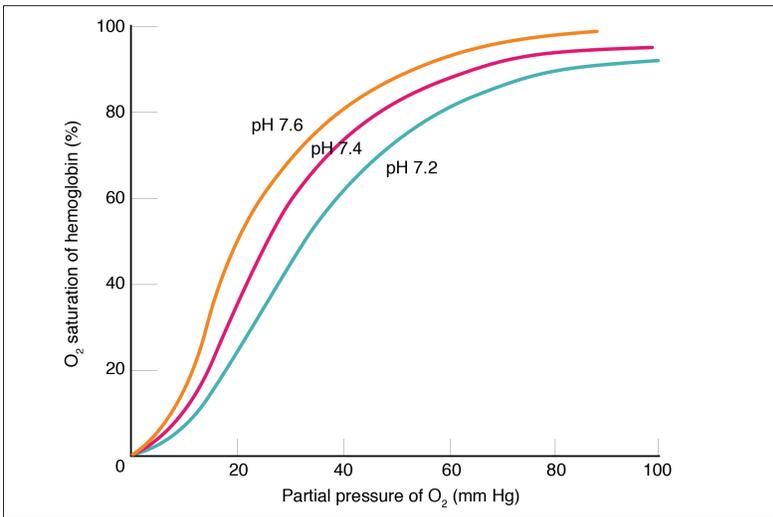
(c) Effect of temperature
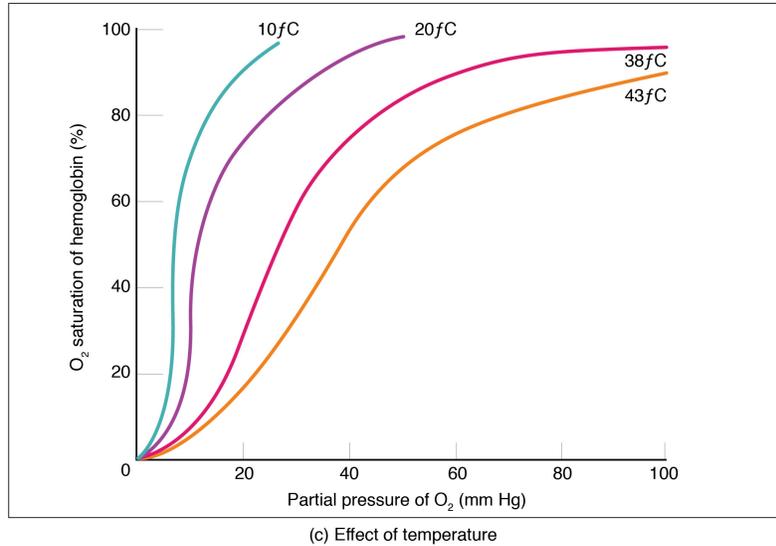
Figure 3.5. Oxygen-hemoglobin dissociation and effects of pH and temperature. These three graphs show (a) the relationship between the partial pressure of oxygen and hemoglobin saturation, (b) the effect of pH on the oxygen-hemoglobin dissociation curve, and (c) the effect of temperature on the oxygen-hemoglobin dissociation curve.

The mechanisms behind the oxygen-hemoglobin saturation/dissociation curve also serve as automatic control mechanisms that regulate how much oxygen is delivered to different tissues throughout the body. This is important because some tissues have a higher metabolic rate than others. Highly active tissues, such as muscle, rapidly use oxygen to produce ATP, lowering the partial pressure of oxygen in the tissue to about 20 mm Hg. The partial pressure of oxygen inside capillaries is about 100 mm Hg, so the difference between the two becomes quite high, about 80 mm Hg. As a result, a greater number of oxygen molecules dissociate from hemoglobin and enter the tissues. The reverse is true of tissues, such as adipose (body fat), which have lower metabolic rates. Because less oxygen is used by these cells, the partial pressure of oxygen within such tissues remains relatively high, resulting in fewer oxygen molecules dissociating from hemoglobin and entering the tissue interstitial fluid. Although venous blood is said to be deoxygenated, some oxygen is still bound to hemoglobin in its red blood cells. This provides an oxygen reserve that can be used when tissues suddenly demand more oxygen.

Factors other than partial pressure also affect the oxygen-hemoglobin saturation/dissociation curve. For example, a higher temperature promotes hemoglobin and oxygen to dissociate faster, whereas a lower temperature inhibits dissociation (see [Figure 3.5b](#)). However, the human body tightly regulates temperature, so this factor may not affect gas exchange throughout the body. The exception to this is in highly active tissues, which may release a larger amount of energy than is given off as heat. As a result, oxygen readily dissociates from hemoglobin, which is a mechanism that helps to provide active tissues with more oxygen.

Certain hormones, such as androgens, epinephrine, thyroid hormones, and growth hormone, can affect the oxygen-hemoglobin saturation/disassociation curve by stimulating the production of a compound called 2,3-bisphosphoglycerate (BPG) by erythrocytes. BPG is a byproduct of glycolysis. Because erythrocytes do not contain mitochondria, glycolysis is the sole method by which these cells produce ATP. BPG promotes the disassociation of oxygen from hemoglobin. Therefore, the greater the concentration of BPG, the more readily oxygen dissociates from hemoglobin, despite its partial pressure.

The pH of the blood is another factor that influences the oxygen-hemoglobin saturation/dissociation curve (see Figure 3.5b). The Bohr effect is a phenomenon that arises from the relationship between pH and oxygen's affinity for hemoglobin: A lower, more acidic pH promotes oxygen dissociation from hemoglobin. In contrast, a higher, or more basic, pH inhibits oxygen dissociation from hemoglobin. The greater the amount of carbon dioxide in the blood, the more molecules that must be converted, which in turn generates hydrogen ions and thus lowers blood pH. Furthermore, blood pH may become more acidic when certain byproducts of cell metabolism, such as lactic acid, carbonic acid, and carbon dioxide, are released into the bloodstream.

Carbon dioxide transport in the blood

Carbon dioxide is transported by three major mechanisms. The first mechanism of carbon dioxide transport is by blood plasma, as some carbon dioxide molecules dissolve in the blood. The second mechanism is transport in the form of bicarbonate (HCO_3^-), which also dissolves in plasma. The third mechanism of carbon dioxide transport is similar to the transport of oxygen by erythrocytes (Figure 3.6).
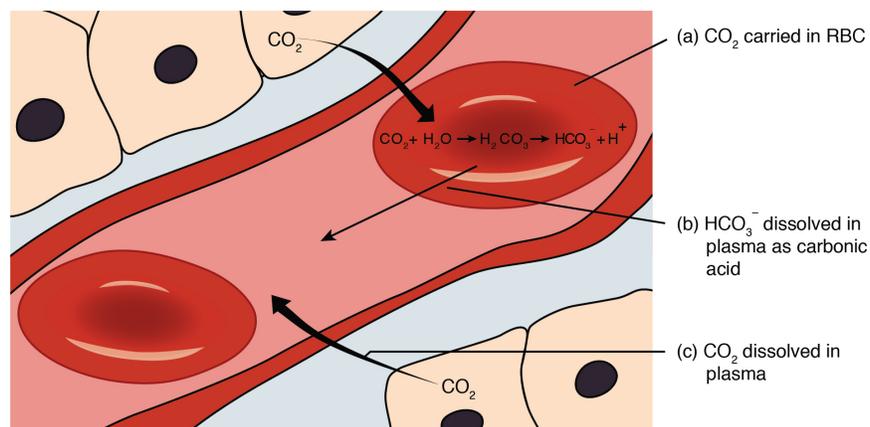

Figure 3.6. Carbon dioxide transport. Carbon dioxide is transported by three different methods: (a) in erythrocytes; (b) after forming carbonic acid ($H2CO3$), which is dissolved in plasma; (c) and in plasma.

Dissolved carbon dioxide

Although carbon dioxide is not considered to be highly soluble in blood, a small fraction—about 7 to 10 percent—of the carbon dioxide that diffuses into the blood from the tissues dissolves in plasma. The dissolved carbon dioxide then travels in the bloodstream and when the blood reaches the pulmonary capillaries, the dissolved carbon dioxide diffuses across the respiratory membrane into the alveoli, where it is then exhaled during pulmonary ventilation.

Bicarbonate buffer

A large fraction—about 70 percent—of the carbon dioxide molecules that diffuse into the blood is transported to the lungs as bicarbonate. Most bicarbonate is produced in erythrocytes after carbon dioxide diffuses into the capillaries, and subsequently into red blood cells. Carbonic anhydrase (CA) causes carbon dioxide and water to form carbonic acid (H_2CO_3), which dissociates into two ions: bicarbonate (HCO_3^-) and hydrogen (H^+). The following formula depicts this reaction:
$$CO2 + H2O \: CA \leftrightarrow H2CO3 \leftrightarrow H+ + HCO3 - CO2 + H2O \: CA \leftrightarrow H2CO3 \leftrightarrow H+ + HCO3 -$$

Bicarbonate tends to build up in the erythrocytes so that there is a greater concentration of bicarbonate in the erythrocytes than in the surrounding blood plasma. As a result, some of the bicarbonate will leave the erythrocytes and move down its concentration gradient into the plasma in exchange for chloride (Cl^-) ions. This phenomenon is referred to as the chloride shift and occurs because by exchanging one negative ion for another negative ion, neither the electrical charge of the erythrocytes nor that of the blood is altered.

At the pulmonary capillaries, the chemical reaction that produced bicarbonate (shown above) is reversed, and carbon dioxide and water are the products. Much of the bicarbonate in the plasma re-enters the erythrocytes in exchange for chloride ions. Hydrogen ions and bicarbonate ions join to form carbonic acid, which is converted into carbon dioxide and water by carbonic anhydrase. Carbon dioxide diffuses out of the erythrocytes and into the plasma, where it can further diffuse across the respiratory membrane into the alveoli to be exhaled during pulmonary ventilation.

Carbaminohemoglobin

About 20 percent of carbon dioxide is bound by hemoglobin and is transported to the lungs. Carbon dioxide does not bind to iron as oxygen does; instead, carbon dioxide binds amino acid moieties on the globin portions of hemoglobin to form carbaminohemoglobin, which forms when hemoglobin and carbon dioxide bind. When hemoglobin is not transporting oxygen, it tends to have a bluish-purple

tone to it, creating the darker maroon color typical of deoxygenated blood. The following formula depicts this reversible reaction:

$CO_2 + Hb \leftrightarrow HbCO_2$

Similar to the transport of oxygen by heme, the binding and dissociation of carbon dioxide to and from hemoglobin is dependent on the partial pressure of carbon dioxide. Because carbon dioxide is released from the lungs, blood that leaves the lungs and reaches body tissues has a lower partial pressure of carbon dioxide than is found in the tissues. As a result, carbon dioxide leaves the tissues because of its higher partial pressure, enters the blood, and then moves into red blood cells, binding to hemoglobin. In contrast, in the pulmonary capillaries, the partial pressure of carbon dioxide is high compared to within the alveoli. As a result, carbon dioxide dissociates readily from hemoglobin and diffuses across the respiratory membrane into the air.

Question 3.1

When ventilation is not sufficient, which of the following occurs?
a. The capillary constricts.
b. The capillary dilates.
c. The partial pressure of oxygen in the affected alveolus increases.
d. The bronchioles dilate.

Question 3.2

Gas exchange that occurs at the level of the tissues is called _____.
a. external respiration
b. interpulmonary respiration
c. internal respiration
d. pulmonary ventilation

Watch this video to see the transport of oxygen from the lungs to the tissues.

Question 3.3

Oxyhemoglobin forms by a chemical reaction between which of the following?
a. hemoglobin and carbon dioxide
b. carbonic anhydrase and carbon dioxide
c. hemoglobin and oxygen
d. carbonic anhydrase and oxygen

Question 3.4

Which of the following factors play a role in the oxygen-hemoglobin saturation/dissociation cu
a. temperature
b. pH
c. BPG
d. All of the above.

Question 3.5

Why is oxygenated blood bright red, whereas deoxygenated blood tends to be more of a purple
the video about oxygen transport).

Question 3.6

Describe the relationship between the partial pressure of oxygen and the binding of oxygen to h

3.2 – Circulatory Systems

The content of this chapter was adapted from the *Concepts of Biology-1st Canadian Edition* open textbook by Charles Molnar and Jane Gair ([Chapter 11.3- Circulatory and Respiratory Systems](#)).

3.2. Compare and contrast different circulatory systems u

sing specific animal examples and evolution.

The circulatory system

The circulatory system is a network of vessels—the arteries, veins, and capillaries—and a pump, the heart. In all vertebrate organisms this is a closed-loop system, in which the blood is largely separated from the body's other extracellular fluid compartment, the interstitial fluid, which is the fluid bathing the cells. Blood circulates inside blood vessels and circulates unidirectionally from the heart around one of two circulatory routes, then returns to the heart again; this is a closed circulatory system. Open circulatory systems are found in invertebrate animals in which the circulatory fluid bathes the internal organs directly even though it may be moved about with a pumping heart.

The heart

The heart is a complex muscle that consists of two pumps: one that pumps blood through pulmonary circulation to the lungs, and the other that pumps blood through systemic circulation to the rest of the body's tissues (and the heart itself).

The heart is asymmetrical, with the left side being larger than the right side, correlating with the different sizes of the pulmonary and systemic circuits ([Figure 3.7](#)). In humans, the heart is about the size of a clenched fist; it is divided into four chambers: two atria and two ventricles. There is one atrium and one ventricle on the right side and one atrium and one ventricle on the left side. The right atrium receives deoxygenated blood from the systemic circulation through the major veins: the

superior vena cava, which drains blood from the head and from the veins that come from the arms, as well as the inferior vena cava, which drains blood from the veins that come from the lower organs and the legs. This deoxygenated blood then passes to the right ventricle through the tricuspid valve, which prevents the backflow of blood. After it is filled, the right ventricle contracts, pumping the blood to the lungs for reoxygenation. The left atrium receives the oxygen-rich blood from the lungs. This blood passes through the bicuspid valve to the left ventricle where the blood is pumped into the aorta. The aorta is the major artery of the body, taking oxygenated blood to the organs and muscles of the body. This pattern of pumping is referred to as double circulation and is found in all mammals. ([Figure 3.7](#)).
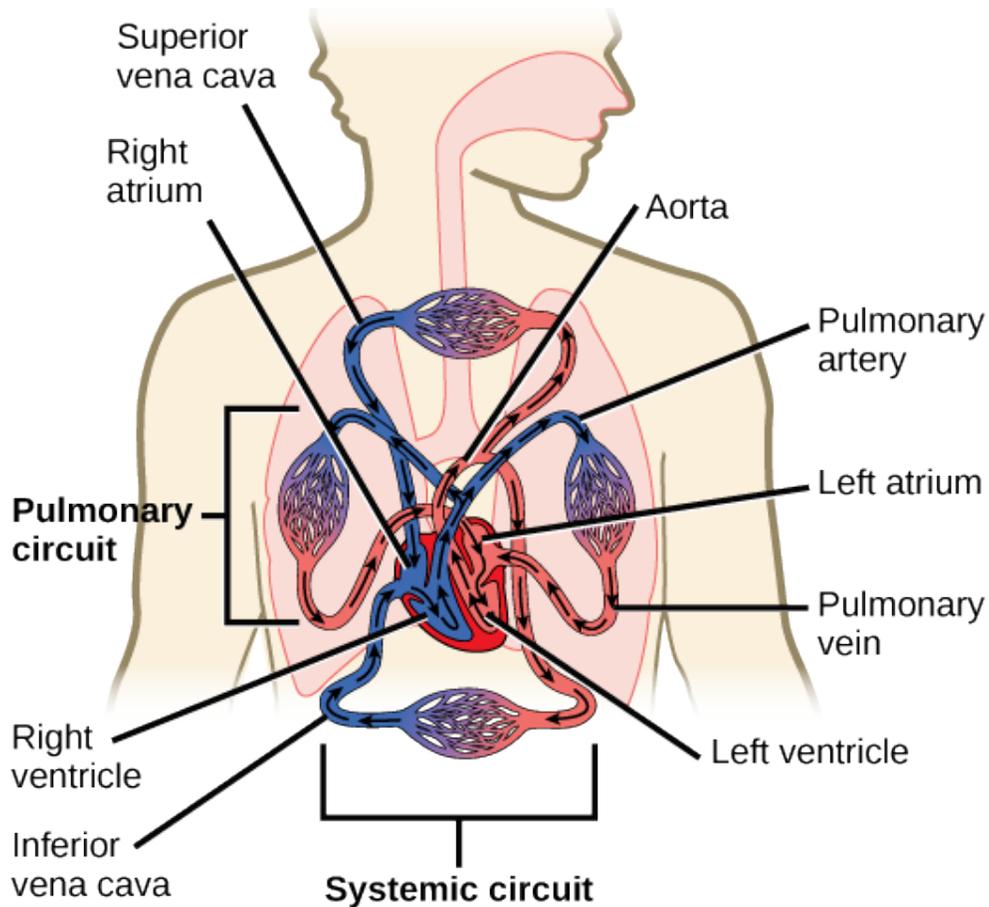

Figure 3.7. The heart is divided into four chambers, two atria, and two ventricles. Each chamber is separated by one-way valves. The right side of the heart receives deoxygenated blood from the body and pumps it to the lungs. The left side of the heart pumps blood to the rest of the body.

Question 3.6

Which of the following statements about the circulatory system is false?
a. Blood in the pulmonary vein is deoxygenated.
b. Blood in the inferior vena cava is deoxygenated.
c. Blood in the pulmonary artery is deoxygenated.
d. Blood in the aorta is oxygenated.

Blood vessels

The blood from the heart is carried through the body by a complex network of blood vessels (Figure 3.8). Arteries take blood away from the heart. The main artery of the systemic circulation is the aorta; it branches into major arteries that take blood to different limbs and organs. The aorta and arteries near the heart have heavy but elastic walls that respond to and smooth out the pressure differences caused by the beating heart. Arteries farther away from the heart have more muscle tissue in their walls that can constrict to affect flow rates of blood. The major arteries diverge into minor arteries, and then smaller vessels called arterioles, to reach more deeply into the muscles and organs of the body.

Arterioles diverge into capillary beds. Capillary beds contain a large number, 10's to 100's of capillaries that branch among the cells of the body. Capillaries are narrow-diameter tubes that can fit single red blood cells and are the sites for the exchange of nutrients, waste, and oxygen with tissues at the cellular level. Fluid also leaks from the blood into the interstitial space from the capillaries. The capillaries converge again into venules that connect to minor veins that finally connect to major veins. Veins are blood vessels that bring blood high in carbon dioxide back to the heart. Veins are not as thick-walled as arteries, since pressure is lower, and they have valves along their length that prevent backflow of blood away from the heart. The major veins drain blood from the same organs and limbs that the major arteries supply.
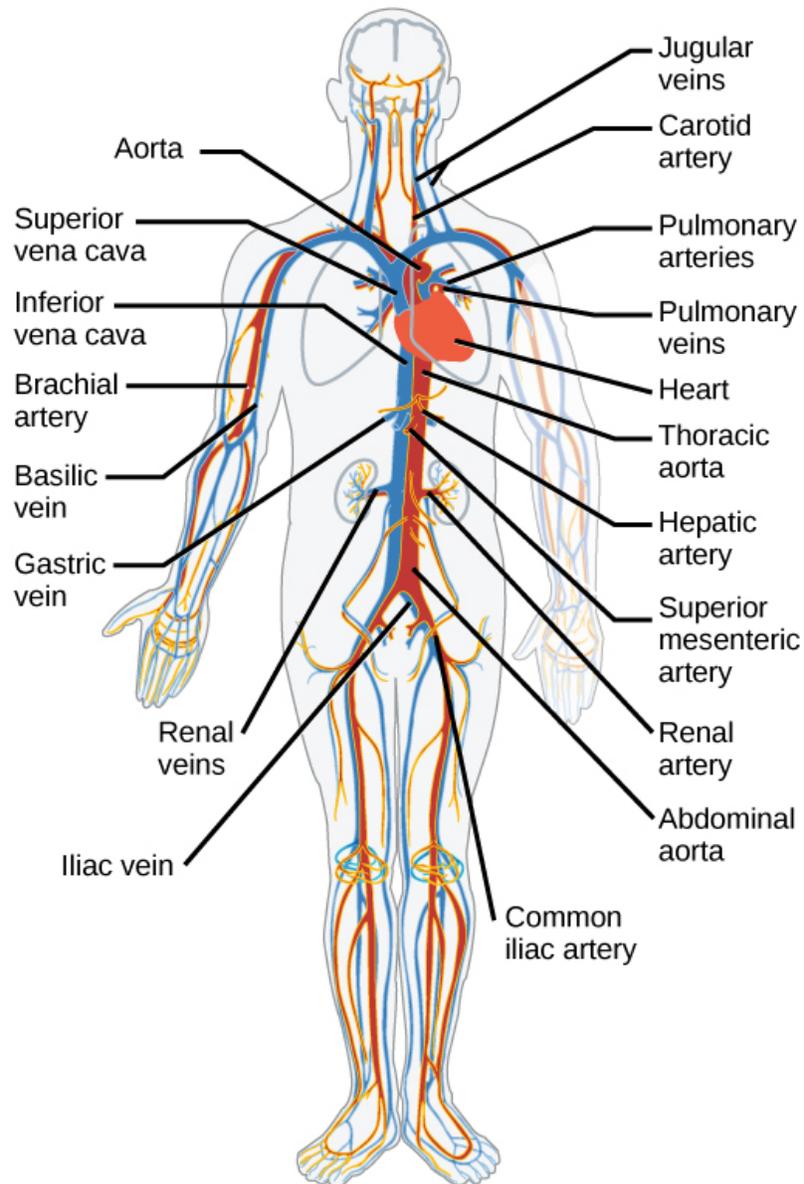
Figure 3.8. The arteries of the body, indicated in red, start at the aortic arch and branch to supply the organs and muscles of the body with oxygenated blood. The veins of the body, indicated in blue, return blood to the heart. The pulmonary arteries are blue to reflect the fact that they are deoxygenated, and the pulmonary veins are red to reflect that they are oxygenated. (credit: modification of work by Mariana Ruiz Villareal)

Blood flow refers to the movement of blood through a vessel, tissue, or organ, and is usually expressed in terms of volume of blood per unit of time. It is initiated by the contraction of the ventricles of the heart. Ventricular contraction ejects blood into the major arteries, resulting in flow from regions of higher pressure to regions of lower pressure, as blood encounters smaller arteries and arterioles, then capillaries, then the venules and veins of the venous system. This section

discusses a number of critical variables that contribute to blood flow throughout the body. It also discusses the factors that impede or slow blood flow, a phenomenon known as resistance.

Hydrostatic pressure is the force exerted by a fluid due to gravitational pull, usually against the wall of the container in which it is located. One form of hydrostatic pressure is blood pressure, the force exerted by blood upon the walls of the blood vessels or the chambers of the heart. Blood pressure may be measured in capillaries and veins, as well as the vessels of the pulmonary circulation; however, the term blood pressure without any specific descriptors typically refers to systemic arterial blood pressure–that is, the pressure of blood flowing in the arteries of the systemic circulation. In clinical practice, this pressure is measured in mm Hg and is usually obtained using the brachial artery of the arm.

The roles of vessel diameter and total area in blood flow and blood pressure

Recall that we classified arterioles as resistance vessels, because given their small lumen, they dramatically slow the flow of blood from arteries. In fact, arterioles are the site of greatest resistance in the entire vascular network. This may seem surprising, given that capillaries have a smaller size. How can this phenomenon be explained?

Figure 3.9 compares vessel diameter, total cross-sectional area, average blood pressure, and blood velocity through the systemic vessels. Notice in parts (a) and (b) that the total cross-sectional area of the body's capillary beds is far greater than any other type of vessel. Although the diameter of an individual capillary is significantly smaller than the diameter of an arteriole, there are vastly more capillaries in the body than there are other types of blood vessels. Part (c) shows that blood pressure drops unevenly as blood travels from arteries to arterioles, capillaries, venules, and veins, and encounters greater resistance. However, the site of the most precipitous drop, and the site of greatest resistance, is the arterioles. This explains why vasodilation and vasoconstriction of arterioles play more significant roles in regulating blood pressure than do the vasodilation and vasoconstriction of other vessels.

Part (d) shows that the velocity (speed) of blood flow decreases dramatically as the blood moves from arteries to arterioles to capillaries. This slow flow rate allows more time for exchange processes to occur. As blood flows through the veins, the rate of velocity increases, as blood is returned to the heart.
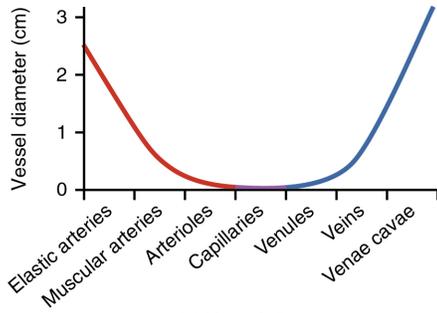
(a) Vessel diameter
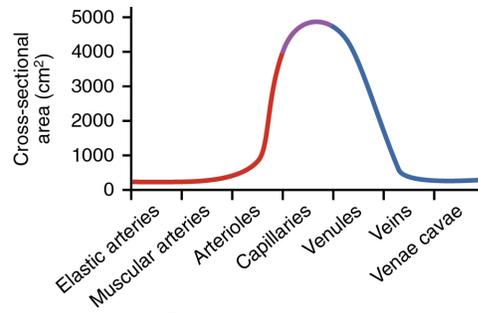
(b) Total cross-sectional area of vessels
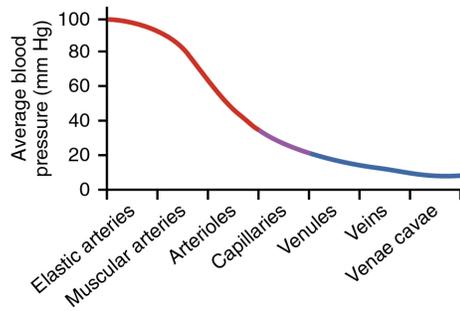
(c) Average blood pressure
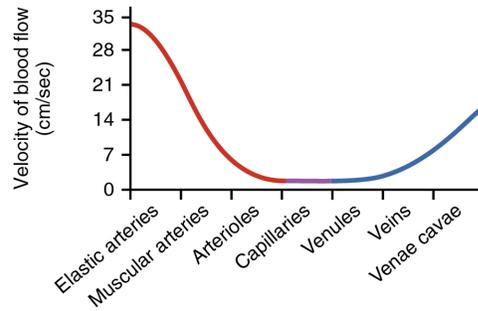
(d) Velocity of blood flow

Figure 3.9. The relationships among blood vessels that can be compared include (a) vessel diameter, (b) total cross-sectional area, (c) average blood pressure, and (d) velocity of blood flow.

Question 3.7

Where would flow be the fastest, narrow or wide tube? Why?

Question 3.8

How is this relevant for blood transport through a vertebrate body?

Question 3.9

How does the giraffe pump blood to its head?

CHAPTER 4 – ION AND WATER BALANCE

4.1. Discuss osmoregulatory function challenges of animals living in terrestrial versus aquatic environments;

4.2. Explain how ion and water balance function in sample animal saltwater systems;

4.3. Explain how ion and water balance function in sample animal freshwater systems; and,

4.4. Explain how ion and water balance function in sample animal terrestrial systems.

4.1 – Osmoregulation in Animals Living in Aquatic Environment

4.1 Discuss osmoregulatory function challenges of animals living in terrestrial versus aquatic environments;

4.2. Explain how ion and water balance function in sample animal saltwater systems; and,

4.3. Explain how ion and water balance function in sample animal freshwater systems.

The content of this chapter was adapted from the *Concepts of Biology-1st Canadian Edition* open textbook by Charles Molnar and Jane Gair ([Chapter 11 – Homeostasis and Osmoregulation](#)).

Osmoregulation

Osmoregulation is the process of maintaining salt and water balance (osmotic balance) across membranes within the body. The fluids inside and surrounding cells are composed of water, electrolytes, and nonelectrolytes. An electrolyte is a compound that dissociates into ions when dissolved in water. A nonelectrolyte, in contrast, does not dissociate into ions in water. The body's fluids include blood plasma, the fluid that exists within cells, and the interstitial fluid that exists in the spaces between cells and tissues of the body. The membranes of the body (both the membranes around cells and the "membranes" made of cells lining body cavities) are semipermeable membranes. Semipermeable membranes are permeable to certain types of solutes and to water, but typically cell membranes are impermeable to solutes.

The body does not exist in isolation. There is a constant input of water and electrolytes into the system. Excess water, electrolytes, and wastes are transported to the kidneys and excreted, helping to maintain osmotic balance. Insufficient fluid intake results in fluid conservation by the kidneys.

Biological systems constantly interact and exchange water and nutrients with the environment by way of consumption of food and water and through excretion in the form of sweat, urine, and feces. Without a mechanism to regulate osmotic pressure, or when a disease damages this mechanism, there is a tendency to accumulate toxic waste and water, which can have dire consequences.

Mammalian systems have evolved to regulate not only the overall osmotic pressure across membranes, but also specific concentrations of important electrolytes in the three major fluid compartments: blood plasma, interstitial fluid, and intracellular fluid. Since osmotic pressure is regulated by the movement of water across membranes, the volume of the fluid compartments can also change temporarily. Since blood plasma is one of the fluid components, osmotic pressures have a direct bearing on blood pressure.

Excretory system

The human excretory system functions to remove waste from the body through the skin as sweat, the lungs in the form of exhaled carbon dioxide, and through the urinary system in the form of urine. All three of these systems participate in osmoregulation and waste removal. Here we focus on the urinary system, which is comprised of the paired kidneys, the ureter, urinary bladder and urethra (Figure 4.1). The kidneys are a pair of bean-shaped structures that are located just below the liver in the body cavity. Each of the kidneys contains more than a million tiny units called nephrons that filter blood containing the metabolic wastes from cells. All the blood in the human body is filtered about 60 times a day by the kidneys. The nephrons remove wastes, concentrate them, and form urine that is collected in the bladder.

Internally, the kidney has three regions—an outer cortex, a medulla in the middle, and the renal pelvis, which is the expanded end of the ureter. The renal cortex contains the nephrons—the functional unit of the kidney. The renal pelvis collects the urine and leads to the ureter on the outside of the kidney. The ureters are urine-bearing tubes that exit the kidney and empty into the urinary bladder.
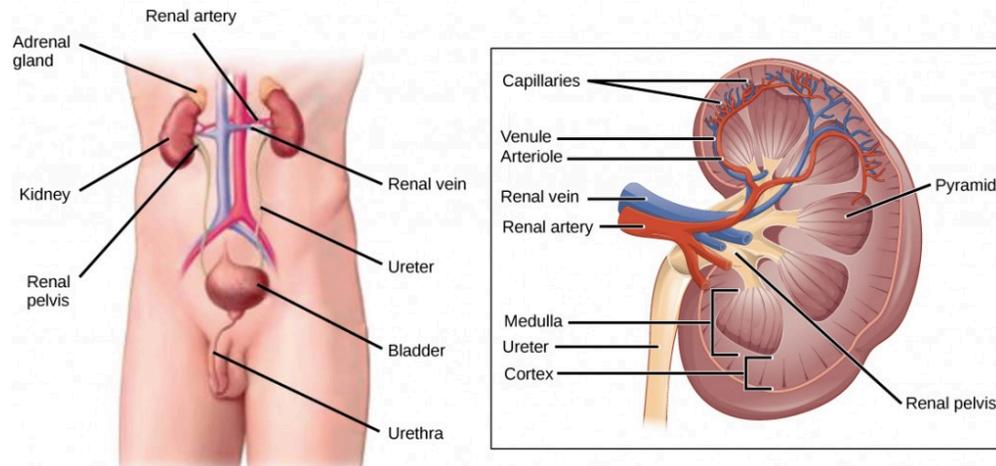
Figure 4.1. The human excretory system is made up of the kidneys, ureter, urinary bladder, and urethra. The kidneys filter blood and form urine, which is stored in the bladder until it is eliminated through the urethra. On the right, the internal structure of the kidney is shown. (credit: modification of work by NCI, NIH)

Blood enters each kidney from the aorta, the main artery supplying the body below the heart, through a renal artery. It is distributed in smaller vessels until it reaches each nephron in capillaries. Within the nephron, the blood comes in intimate contact with the waste-collecting tubules in a structure called the glomerulus. Water and many solutes present in the blood, including ions of sodium, calcium, magnesium, and others; as well as wastes and valuable substances such as amino acids, glucose, and vitamins, leave the blood and enter the tubule system of the nephron. As materials pass through the tubule much of the water, required ions, and useful compounds are reabsorbed back into the capillaries that surround the tubules leaving the wastes behind. Some of this reabsorption requires active transport and consumes ATP. Some wastes, including ions and some drugs remaining in the blood, diffuse out of the capillaries into the interstitial fluid and are taken up by the tubule cells. These wastes are then actively secreted into the tubules. The blood then collects in larger and larger vessels and leaves the kidney in the renal vein. The renal vein joins the inferior vena cava, the main vein that returns blood to the heart from the lower body. The amounts of water and ions reabsorbed into the circulatory system are carefully regulated and this is an important way the body regulates its water content and ion levels. The waste is collected in larger tubules and then leaves the kidney in the ureter, which leads to the bladder where urine, the combination of waste materials and water, is stored.

The bladder contains sensory nerves, stretch receptors that signal when it needs to be emptied. These signals create the urge to urinate, which can be voluntarily suppressed up to a limit. The conscious decision to urinate sets in play signals that open the sphincters, rings of smooth muscle that close off the opening, to the urethra that allows urine to flow out of the bladder and the body.

Osmosis is the diffusion of water across a membrane in response to **osmotic pressure** caused by

an imbalance of molecules on either side of the membrane. **Osmoregulation** is the process of maintenance of salt and water balance (**osmotic balance**) across membranes within the body's fluids, which are composed of water, plus electrolytes and non-electrolytes. An **electrolyte** is a solute that dissociates into ions when dissolved in water. A **non-electrolyte**, in contrast, doesn't dissociate into ions during water dissolution. Both electrolytes and non-electrolytes contribute to the osmotic balance. The body's fluids include blood plasma, the cytosol within cells, and interstitial fluid, the fluid that exists in the spaces between cells and tissues of the body. The membranes of the body (such as the pleural, serous, and cell membranes) are **semi-permeable membranes**. Semi-permeable membranes are permeable (or permissive) to certain types of solutes and water. Solutions on two sides of a semi-permeable membrane tend to equalize in solute concentration by movement of solutes and/or water across the membrane. As seen in Figure 4.2, a cell placed in water tends to swell due to gain of water from the hypotonic or "low salt" environment. A cell placed in a solution with higher salt concentration, on the other hand, tends to make the membrane shrivel up due to loss of water into the hypertonic or "high salt" environment. Isotonic cells have an equal concentration of solutes inside and outside the cell; this equalizes the osmotic pressure on either side of the cell membrane which is a semi-permeable membrane.
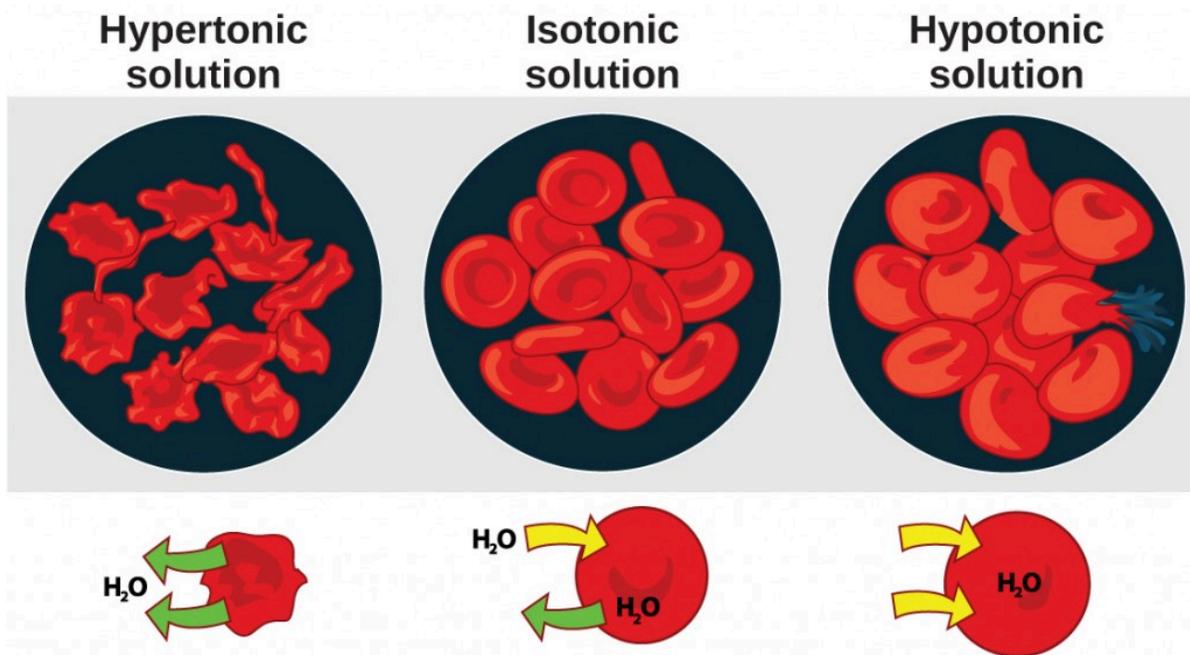
Figure 4.2. Cells placed in a hypertonic environment tend to shrink due to loss of water. In a hypotonic environment, cells tend to swell due to the intake of water. The blood maintains an isotonic environment so that cells neither shrink nor swell. (credit: Mariana Ruiz Villareal)

The body does not exist in isolation. There is a constant input of water and electrolytes into the system. While osmoregulation is achieved across membranes within the body, excess electrolytes and wastes are transported to the kidneys and excreted, helping to maintain osmotic balance.

Need for osmoregulation

Biological systems constantly interact and exchange water and nutrients with the environment by way of consumption of food and water and through excretion in the form of sweat, urine, and feces. Without a mechanism to regulate osmotic pressure, or when a disease damages this mechanism, there is a tendency to accumulate toxic waste and water, which can have dire consequences.

Mammalian systems have evolved to regulate not only the overall osmotic pressure across membranes, but also specific concentrations of important electrolytes in the three major fluid compartments: blood plasma, extracellular fluid, and intracellular fluid. Since osmotic pressure is regulated by the movement of water across membranes, the volume of the fluid compartments can also change temporarily. Because blood plasma is one of the fluid components, osmotic pressures have a direct bearing on blood pressure.

Osmotic pressure is influenced by the concentration of solutes in a solution. It is directly proportional to the number of solute atoms or molecules and not dependent on the size of the solute molecules. Because electrolytes dissociate into their component ions, they, in essence, add more solute particles into the solution and have a greater effect on osmotic pressure, per mass than compounds that do not dissociate in water, such as glucose.

Water can pass through membranes by passive diffusion. If electrolyte ions could passively diffuse across membranes, it would be impossible to maintain specific concentrations of ions in each fluid compartment, therefore, they require special mechanisms to cross the semi-permeable membranes in the body. This movement can be accomplished by facilitated diffusion and active transport. Facilitated diffusion requires protein-based channels for moving the solute. Active transport requires energy in the form of ATP conversion, carrier proteins, or pumps in order to move ions against the concentration gradient.

Osmoregulators and osmoconformers

Persons lost at sea without any fresh water to drink, are at risk of severe dehydration because the human body cannot adapt to drinking seawater, which is hypertonic in comparison to body fluids. Organisms such as goldfish that can tolerate only a relatively narrow range of salinity are referred to as stenohaline. About 90 percent of all bony fish are restricted to either freshwater or seawater. They are incapable of osmotic regulation in the opposite environment. It is possible, however, for a few fishes like salmon to spend part of their life in freshwater and part in sea water. Organisms like the salmon and molly that can tolerate a relatively wide range of salinity are referred to as euryhaline organisms. This is possible because some fish have evolved **osmoregulatory** mechanisms to survive in all kinds of aquatic environments. When they live in fresh water, their bodies tend to take up water because the environment is relatively hypotonic, as illustrated in Figure 4.3a. In such

hypotonic environments, these fish do not drink much water. Instead, they pass a lot of very dilute urine, and they achieve electrolyte balance by active transport of salts through the gills. When they move to a hypertonic marine environment, these fish start drinking sea water; they excrete the excess salts through their gills and their urine, as illustrated in Figure 4.3b. Most marine invertebrates, on the other hand, maybe isotonic with sea water (**osmoconformers**). Their body fluid concentrations conform to changes in seawater concentration. Cartilaginous fishes' salt composition of the blood is similar to bony fishes; however, the blood of sharks contains the organic compounds urea and trimethylamine oxide (TMAO). This does not mean that their electrolyte composition is similar to that of seawater. They achieve isotonicity with the sea by storing large concentrations of urea. These animals that secrete urea are called ureotelic animals. TMAO stabilizes proteins in the presence of high urea levels, preventing the disruption of peptide bonds that would occur in other animals exposed to similar levels of urea. Sharks are cartilaginous fish with a rectal gland to secrete salt and assist in osmoregulation.

Absorbs water through skin

Actively takes up ions through gills

Drinks little water

Movement of water
Movement of ions

Excretes dilute urine

(a) Osmoregulation in a freshwater environment

Loses water through skin

Drinks ample water

Direction of ion movement (Na^+,K^+,Cl^-)

Direction of water movement

Excretes ions through gills

Excretes concentrated urine

(b) Osmoregulation in a saltwater environment

Figure 4.3. Fish are osmoregulators, but must use different mechanisms to survive in (a) freshwater or (b) saltwater environments. (credit: modification of work by Duane Raver, NOAA)

Different osmoregulatory challenges and solutions in aquatic and terrestrial environments

Fish living in freshwater requirements have very different challenges in terms of ion and water balance in their body than the fish living in saltwater environments. Although structurally some solutions to these challenges are similar (chloride cells in gills and nephrons), the functional aspect of each solution is different. The next two figures (Figure 4.4 and Figure 4.5) offer a summary of these solutions and introduce challenges in solutions on land as well, which are discussed in further details in Chapter section 4.2.

A. Mammals	B. Freshwater Fish	C. Marine Fish	D. Insects

i) Movement of salt and water through the organism

ii) Structure of the organisms functional unit

LEGEND

Part i)
- ← Water
- ← Salt

Part ii)
- GL - Glomerulus
- BC - Bowman's Capsule
- PT - Proximal Convoluted Tubule
- DT - Distal Convoluted Tubule
- CT - Collecting Tubule
- LH - Loop of Henle
- CD - Collecting Duct
- IS - Intermediate Segment
- RC - Rectal Canal
- HG - Hindgut
- MT - Malpighian Tubules
- HM - Hemolymph

figure 1: Comparing the movement of salt and water through different animals at different levels of magnification.

Figure 4.4. Comparison of movement of water and salt through different animals. Figure created and designed by Emily Taylor, a graduate student in Biomedical Communications Masters program at the University of Toronto Mississauga (developed as a project in MSC2001Y Visual Representation of Medical Knowledge course in the BMC MSc program). Please note that in B. freshwater fish should be referred to as hypertonic to their environment (or as in living in hypotonic environment). In C, the opposite should be the case.

A. Mammals:
i) Mammals intake water and salts through their diets and drinking. The urine creation and reabsorption occurs in the mammalian kidneys. Excretion is through the bladder and urethra.
ii) The mammalian nephron has a glomerulus, proximal tubule, loop of Henle and distal tubule which feed into the collecting duct. The blood in the capillary bed moves countercurrent to filtrate flow to promote re-absorption.

B. Fresh Water Fish:
i) Fresh water (FW) fish are hypotonic to their environment and so limit water intake. They do not ingest water by drinking but obtain it and salt through diet. The gills are used to intake more salt and water. Their urine is hypoosmotic to their body.
ii) The FW fish nephron does not contain a loop of Henle but does have an intermediate segment. The collecting ducts are impermeable to water to prevent water from reabsorbing into the blood.

C. Marine Fish:
i) Marine fish are hypertonic to their environment and consume water to prevent excessive loss. They ingest water and salt through drink and diet. Their gills moderate the levels in the body, by removing both. The fish conserve water by producing concentrated urine in the kidney.
ii) The nephron of marine fish contain a smaller glomerulus, or none at all. They do not have a distal tubule and have a reduced proximal tubule. The collecting duct is permeable to water to allow movement back into the blood.

D. Insects:
i) The insect obtains salts and water through diet and drinking. Insects are able to regulate the concentration of the excrement, making it hyperosmotic to the body if necessary. It is excreted out of the hindgut.
ii) In the malpighian tubules the primary urine is created. The tubules deposit primary urine into the midgut. In the rectum, most of the ions and water are reabsorbed.
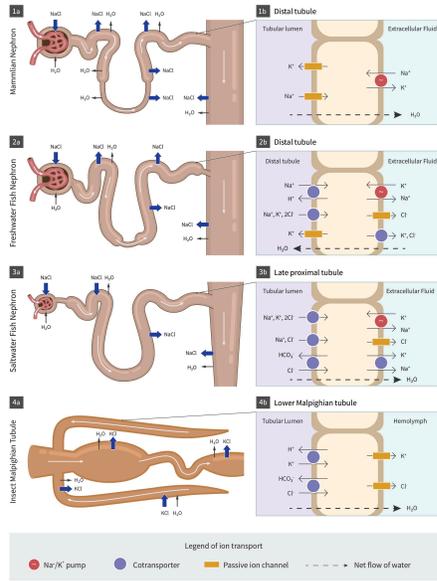
Figure XX. Comparing salt and water movement of mammalian, freshwater fish, saltwater fish nephron, and insect urinary system.

A generalised water and salt movement is depicted for the four organisms corresponding to their relative hyper or hypo-osmotic environment. A schematic depiction of major active and passive transport for each organism in the distal epithelial membrane are represented. For the anatomy of the nephron, see corresponding Figure XX.

Mammalian Nephron

1a. The mammalian nephron evolved to reabsorb water and salt. Filtrate first moves into the glomerulus and water is absorbed into the blood through the proximal tubule, descending limb, distal tubule, and collecting duct. Salts are reabsorbed in the proximal tubule, ascending loop, distal tubule, and collecting duct.

1b. The distal tubule of the mammary nephron contains active Na$^+$/K$^+$ ATPase pumps and passive Na$^+$ and K$^+$ channels that direct the reabsorption of salt ions and water.

Freshwater Fish Nephron

2a. Freshwater teleost (FWT) nephron evolved as a water excreting and salt reabsorbing unit. Salt is actively reabsorbed through the proximal segment, distal tubule, and collecting duct of the nephron. FWT has an enlarged glomerulus and lacks a nephron loop to conserve salt as well as to excrete water. As a result, FWT produce dilute urine in large volumes.

2b. In the distal tubule, salt ions are actively reabsorbed through Na$^+$/K$^+$ ATPase which creates an electrochemical gradient (Na$^+$, K$^+$, and Cl$^-$ ions follow this gradient). K$^+$/Cl$^-$ and Na$^+$/K$^+$/2Cl$^-$ cotransporters reabsorb Cl$^-$ ions. Furthermore, Na$^+$ is absorbed by Na$^+$/H$^+$ exchanger.

Saltwater Fish Nephron

3a. Saltwater teleost (SWT) evolved to conserve water and to excrete salt to maintain osmotic homeostasis and often displays a small (or lacks a) glomerulus. The SWT may also lack a distal tubule and has a reduced proximal tubule. Reabsorption of water occurs mainly in the proximal tubule and collecting duct. Tubular secretion of salt mainly occurs in the early proximal tubule through the presence of active pumps.

3b. The late proximal tubule involves fluid reabsorption driven by active Na$^+$/K$^+$ ATPase which creates a Na$^+$ electrochemical gradient back into the blood (Na$^+$, K$^+$, and Cl$^-$ ions follow this gradient). Na$^+$/Cl$^-$ symporters and the Na$^+$/K$^+$/2Cl$^-$ cotransporters are responsible for Na$^+$ and some Cl$^-$ reabsorption. Further reabsorption of Cl$^-$ occurs by Cl$^-$ channels and through exchange for HCO3$^-$. The transporters allow for a net reabsorption of water.

Insect Urinary System

4a. Water and salt are first filtered through the Malpighian tubules and direct its way to the hindgut and rectum. See previous Figure XX for reference. Salt and water are reabsorbed predominantly in the hindgut and rectum.

4b. The proximal segment of the Malpighian tubules has H$^+$/K$^+$ ATPase transporter and K+ channels that reabsorb K$^+$ ions. Cl$^-$ ions are reabsorbed through Cl$^-$/HCO3$^-$ exchanger and channels.

Figure 4.5. Comparison of nephron systems and channels in animals. This figure was created and designed by Su Min Suh, a graduate student in Biomedical Communications Masters program at the University of Toronto Mississauga (developed as a project in MSC2001Y Visual Representation of Medical Knowledge course in the BMC MSc program).
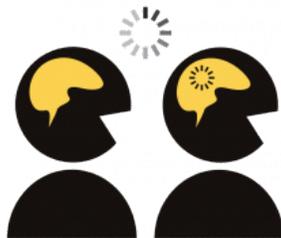
Question 4.1

What is the difference between osmoregulator and osmoconformer?

Question 4.2

What are the biggest osmoregulatory challenges for fish in saltwater?
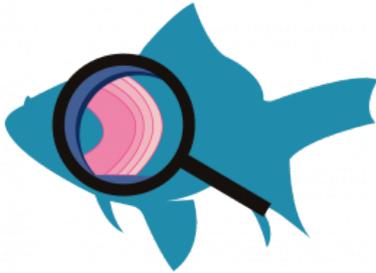
Question 4.3

Summarize how salmon survive in both salt and freshwater. Use this journal article to learn about salmon migration in more details.

4.2 – Ion and Water Balance on Land

4.4. Explain how ion and water balance function in sample animal terrestrial systems.

The content of this chapter was adapted from the *Concepts of Biology-1st Canadian Edition* open textbook by Charles Molnar and Jane Gair ([Chapter 22 – Kidneys and Osmoregulation](#)).

Although the kidneys are the major osmoregulatory organ, the skin and lungs also play a role in the process. Water and electrolytes are lost through sweat glands in the skin, which helps moisturize and cool the skin surface, while the lungs expel a small amount of water in the form of mucous secretions and via evaporation of water vapor.

Kidneys: the main osmoregulatory organ

The kidneys, illustrated in Figure 4.6, are a pair of bean-shaped structures that are located just below and posterior to the liver in the peritoneal cavity. The adrenal glands sit on top of each kidney and are also called the suprarenal glands. Kidneys filter blood and purify it. All the blood in the human body is filtered many times a day by the kidneys; these organs use up almost 25 percent of the oxygen absorbed through the lungs to perform this function. Oxygen allows the kidney cells to efficiently manufacture chemical energy in the form of ATP through aerobic respiration. The filtrate coming out of the kidneys is called urine.
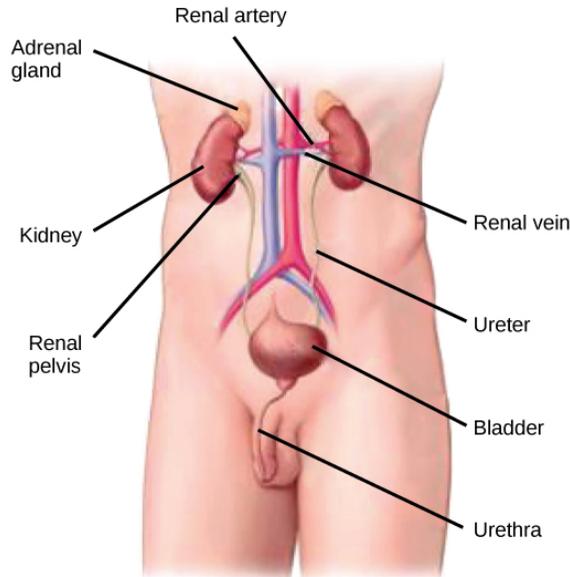
Figure 4.6. Kidneys filter the blood, producing urine that is stored in the bladder prior to elimination through the urethra. (credit: modification of work by NCI)

Kidney structure

Externally, the kidneys are surrounded by three layers, illustrated in <u>Figure 4.7</u>. The outermost layer is a tough connective tissue layer called the renal fascia. The second layer is called the perirenal fat capsule, which helps anchor the kidneys in place. The third and innermost layer is the renal capsule. Internally, the kidney has three regions—an outer cortex, a medulla in the middle, and the renal pelvis in the region called the hilum of the kidney. The hilum is the concave part of the bean-shape where blood vessels and nerves enter and exit the kidney; it is also the point of exit for the ureters. The renal cortex is granular due to the presence of nephrons—the functional unit of the kidney. The medulla consists of multiple pyramidal tissue masses, called the renal pyramids. In between the pyramids are spaces called renal columns through which the blood vessels pass. The tips of the pyramids, called renal papillae, point toward the renal pelvis. There are, on average, eight renal pyramids in each kidney. The renal pyramids along with the adjoining cortical region are called the lobes of the kidney. The renal pelvis leads to the ureter on the outside of the kidney. On the inside of the kidney, the renal pelvis branches out into two or three extensions called the major calyces, which further branch into the minor calyces. The ureters are urine-bearing tubes that exit the kidney and empty into the urinary bladder.
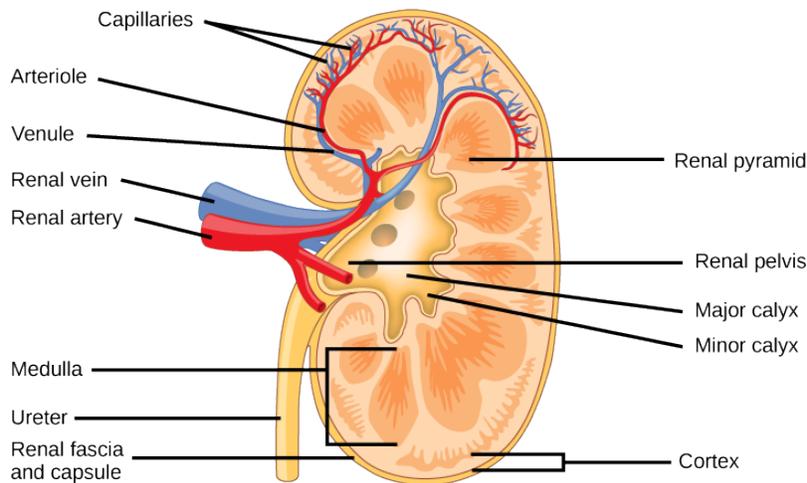
Figure 4.7 The internal structure of the kidney is shown. (credit: modification of work by NCI)

Question 4.4

Which of the following statements about the kidney is false?
a. The renal pelvis drains into the ureter.
b. The renal pyramids are in the medulla.
c. The cortex covers the capsule.
d. Nephrons are in the renal cortex.

Because the kidney filters blood, its network of blood vessels is an important component of its structure and function. The arteries, veins, and nerves that supply the kidney enter and exit at the renal hilum. Renal blood supply starts with the branching of the aorta into the renal arteries(which are each named based on the region of the kidney they pass through) and ends with the exiting of the renal veins to join the inferior vena cava. The renal arteries split into several segmental arteries upon entering the kidneys. Each segmental artery splits further into several interlobar arteries and enters the renal columns, which supply the renal lobes. The interlobar arteries split at the junction of the renal cortex and medulla to form the arcuate arteries. The arcuate "bow-shaped" arteries form arcs along the base of the medullary pyramids. Corticalradiate arteries, as the name suggests, radiate out from the arcuate arteries. The cortical radiate arteries branch into numerous

afferent arterioles and then enter the capillaries supplying the nephrons. Veins trace the path of the arteries and have similar names, except there are no segmental veins.

As mentioned previously, the functional unit of the kidney is the nephron, illustrated in [Figure 4.8](#). Each kidney is made up of over one million nephrons that dot the renal cortex, giving it a granular appearance when sectioned sagittally. There are two types of nephrons– cortical nephrons (85 percent), which are deep in the renal cortex, and juxtamedullary nephrons, which lie in the renal cortex close to the renal medulla. A nephron consists of three parts–a renal corpuscle, a renal tubule, and the associated capillary network, which originates from the cortical radiate arteries.
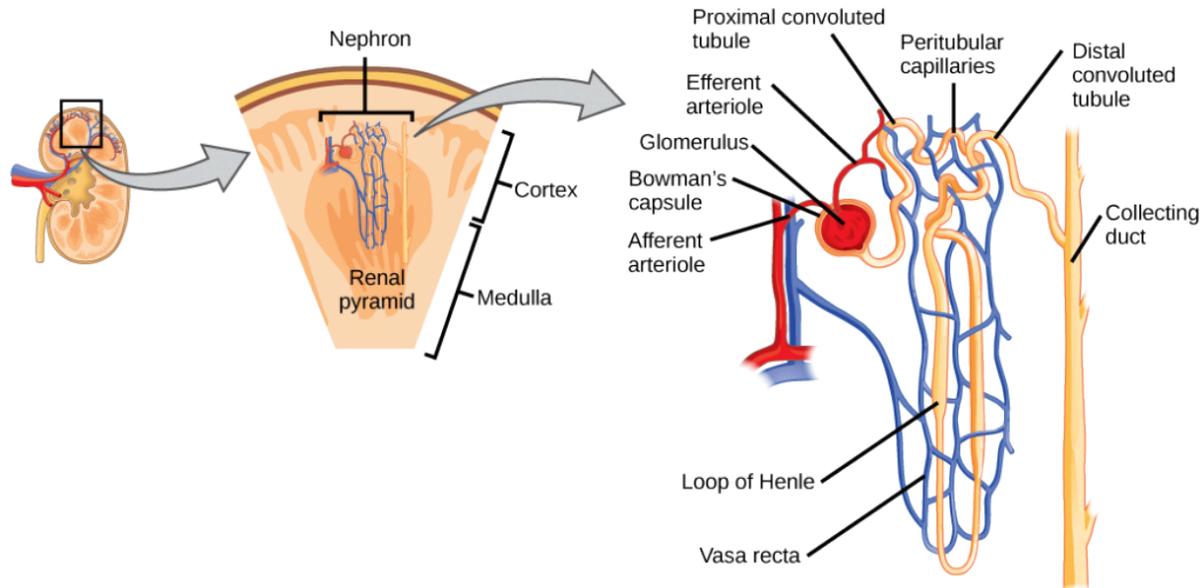
Figure 4.8. The nephron is the functional unit of the kidney. The glomerulus and convoluted tubules are located in the kidney cortex while collecting ducts are located in the pyramids of the medulla. (credit: modification of work by NIDDK)

Nephrons: the functional unit

Nephrons take a simple filtrate of the blood and modify it into the urine. Many changes take place in the different parts of the nephron before the urine is created for disposal. The term forming urine will be used hereafter to describe the filtrate as it is modified into the true urine. The principal task of the nephron population is to balance the plasma to homeostatic set points and excrete potential toxins in the urine. They do this by accomplishing three principle functions–filtration, reabsorption, and secretion. They also have additional secondary functions that exert control in three areas: blood pressure (via production of renin), red blood cell production (via the hormone EPO), and calcium absorption (via conversion of calcidiol into calcitriol, the active form of vitamin D).

Go to [this website](#) to see another coronal section of the kidney and to explore an animation of the workings of nephrons.

Question 4.5

Which of the following statements about the nephron is false?
a. The collecting duct empties into the distal convoluted tubule.
b. The Bowman's capsule surrounds the glomerulus.
c. The loop of Henle is between the proximal and distal convoluted tubules.
d. The loop of Henle empties into the distal convoluted tubule.

Kidney function and physiology

Kidneys filter blood in a three-step process. First, the nephrons filter blood that runs through the capillary network in the glomerulus. Almost all solutes, except for proteins, are filtered out into the glomerulus by a process called glomerular filtration. Second, the filtrate is collected in the renal tubules. Most of the solutes get reabsorbed in the PCT by a process called tubular reabsorption. In the loop of Henle, the filtrate continues to exchange solutes and water with the renal medulla and the peritubular capillary network. Water is also reabsorbed during this step. Then, additional solutes and wastes are secreted into the kidney tubules during tubular secretion, which is, in essence, the opposite process of tubular reabsorption. The collecting ducts collect filtrate coming from the nephrons and fuse in the medullary papillae. From here, the papillae deliver the filtrate, now called urine, into the minor calyces that eventually connect to the ureters through the renal pelvis. This entire process is illustrated in [Figure 4.9](#).

2. Proximal convoluted tubule: reabsorbs ions, water, and nutrients; removes toxins and adjusts filtrate pH

5. Distal tubule: selectively secretes and absorbs different ions to maintain blood pH and electrolyte balance

1. Glomerulus: filters small solutes from the blood

6. Collecting duct: reabsorbs solutes and water from the filtrate

4. Ascending loop of Henle: reabsorbs Na^+ and Cl^- from the filtrate into the interstitial fluid

3. Descending loop of Henle: aquaporins allow water to pass from the filtrate into the interstitial fluid

Figure 4.9. Each part of the nephron performs a different function in filtering waste and maintaining homeostatic balance. (1) The glomerulus forces small solutes out of the blood by pressure. (2) The proximal convoluted tubule reabsorbs ions, water, and nutrients from the filtrate into the interstitial fluid, and actively transports toxins and drugs from the interstitial fluid into the filtrate. The proximal convoluted tubule also adjusts blood pH by selectively secreting ammonia (NH3) into the filtrate, where it reacts with H+ to form NH4+. The more acidic the filtrate, the more ammonia is secreted. (3) The descending loop of Henle is lined with cells containing aquaporins that allow water to pass from the filtrate into the interstitial fluid. (4) In the thin part of the ascending loop of Henle, Na+ and Cl– ions diffuse into the interstitial fluid. In the thick part, these same ions are actively transported into the interstitial fluid. Because salt but not water is lost, the filtrate becomes more dilute as it travels up the limb. (5) In the distal convoluted tubule, K+ and H+ ions are selectively secreted into the filtrate, while Na+, Cl–, and HCO3– ions are reabsorbed to maintain pH and electrolyte balance in the blood. (6) The collecting duct reabsorbs solutes and water from the filtrate, forming dilute urine. (credit: modification of work by NIDDK)

Tubular reabsorption and secretion

Tubular reabsorption occurs in the PCT part of the renal tubule. Almost all nutrients are reabsorbed, and this occurs either by passive or active transport. Reabsorption of water and some key electrolytes are regulated and can be influenced by hormones. Sodium (Na+) is the most abundant ion and most of it is reabsorbed by active transport and then transported to the peritubular capillaries. Because Na+ is actively transported out of the tubule, water follows it to even out the osmotic pressure. Water is also independently reabsorbed into the peritubular capillaries due to the presence of aquaporins, or water channels, in the PCT. This occurs due to the low blood pressure and high osmotic pressure in the peritubular capillaries. However, every solute has a transport maximum and the excess is not reabsorbed.

In the loop of Henle, the permeability of the membrane changes. The descending limb is permeable to water, not solutes; the opposite is true for the ascending limb. Additionally, the loop of Henle invades the renal medulla, which is naturally high in salt concentration and tends to absorb water from the renal tubule and concentrate the filtrate. The osmotic gradient increases as it moves deeper into the medulla. Because two sides of the loop of Henle perform opposing functions, as illustrated in Figure 4.10, it acts as a countercurrent multiplier. The vasa recta around it acts as the countercurrent exchanger.
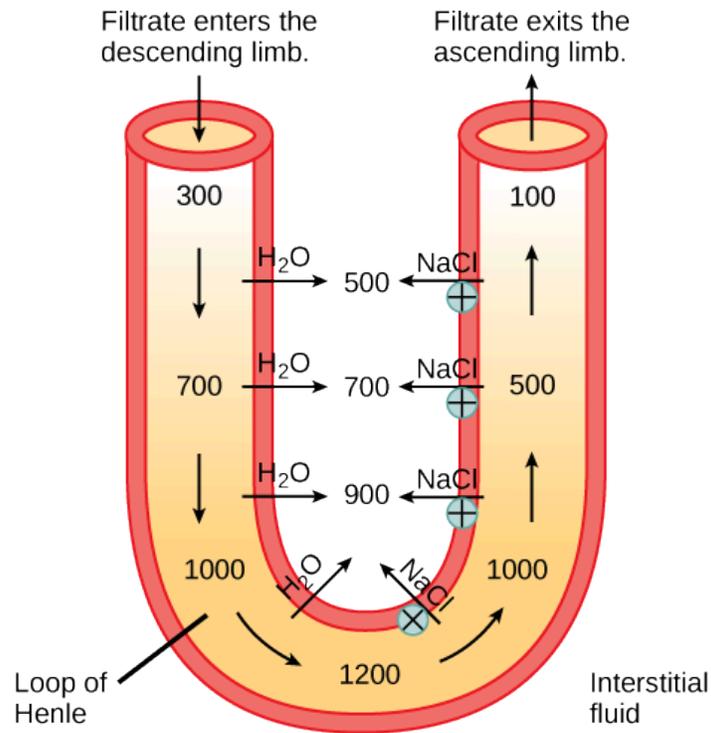
Figure 4.10. The loop of Henle acts as a countercurrent multiplier that uses energy to create concentration gradients. The descending limb is water permeable. Water flows from the filtrate to the interstitial fluid, so osmolality inside the limb increases as it descends into the renal medulla. At the bottom, the osmolality is higher inside the loop than in the interstitial fluid. Thus, as filtrate enters the ascending limb, Na+ and Cl- ions exit through ion channels present in the cell membrane. Further up, Na+ is actively transported out of the filtrate and Cl- follows. Osmolarity is given in units of milliosmoles per liter (mOsm/L).

By the time the filtrate reaches the DCT, most of the urine and solutes have been reabsorbed. If the body requires additional water, all of it can be reabsorbed at this point. Further reabsorption is controlled by hormones, which will be discussed in a later section. Excretion of wastes occurs due to lack of reabsorption combined with tubular secretion. Undesirable products like metabolic wastes, urea, uric acid, and certain drugs, are excreted by tubular secretion. Most of the tubular secretion happens in the DCT, but some occurs in the early part of the collecting duct. Kidneys also maintain an acid-base balance by secreting excess H+ ions.

Although parts of the renal tubules are named proximal and distal, in a cross-section of the kidney, the tubules are placed close together and in contact with each other and the glomerulus. This allows for the exchange of chemical messengers between the different cell types. For example, the DCT ascending limb of the loop of Henle has masses of cells called macula densa, which are in contact with cells of the afferent arterioles called juxtaglomerular cells. Together, the macula densa and

juxtaglomerular cells form the juxtaglomerular complex (JGC). The JGC is an endocrine structure that secretes the enzyme renin and the hormone erythropoietin. When hormones trigger the macula densa cells in the DCT due to variations in blood volume, blood pressure, or electrolyte balance, these cells can immediately communicate the problem to the capillaries in the afferent and efferent arterioles, which can constrict or relax to change the glomerular filtration rate of the kidneys.

To learn more about the vascular system of kidneys, click through this review and the steps of blood flow.

CHAPTER 5 – DIGESTION AND ENERGY

5.1. List and briefly explain different types of energy relevant in biology;

5.2. Compare and contrast digestive system structures in sample animal systems; and,

5.3. Summarize chemical digestion of carbohydrates, proteins and fats and energy metabolism.

The content of this chapter was adapted from the *Concepts of Biology-1st Canadian Edition* open textbook by Charles Molnar and Jane Gair and *Anatomy and Physiology* open textbook.

5.1 – Biological Energy

5.1. List and briefly explain different types of energy relevant for biology.

Cell's metabolism and energy

Scientists use the term bioenergetics to describe the concept of energy flow ([Figure 5.1](#)) through living systems, such as cells. **Cellular processes** such as the building and breaking down of complex molecules **occur through stepwise chemical reactions**. Some of these chemical reactions are spontaneous and release energy, whereas others require energy to proceed. Just as living things must continually consume food to replenish their energy supplies, cells must continually produce more energy to replenish that used by the many energy-requiring chemical reactions that constantly take place. Together, **all of the chemical reactions** that take place inside cells, including those that consume or generate energy, are referred to as the **cell's metabolism**.
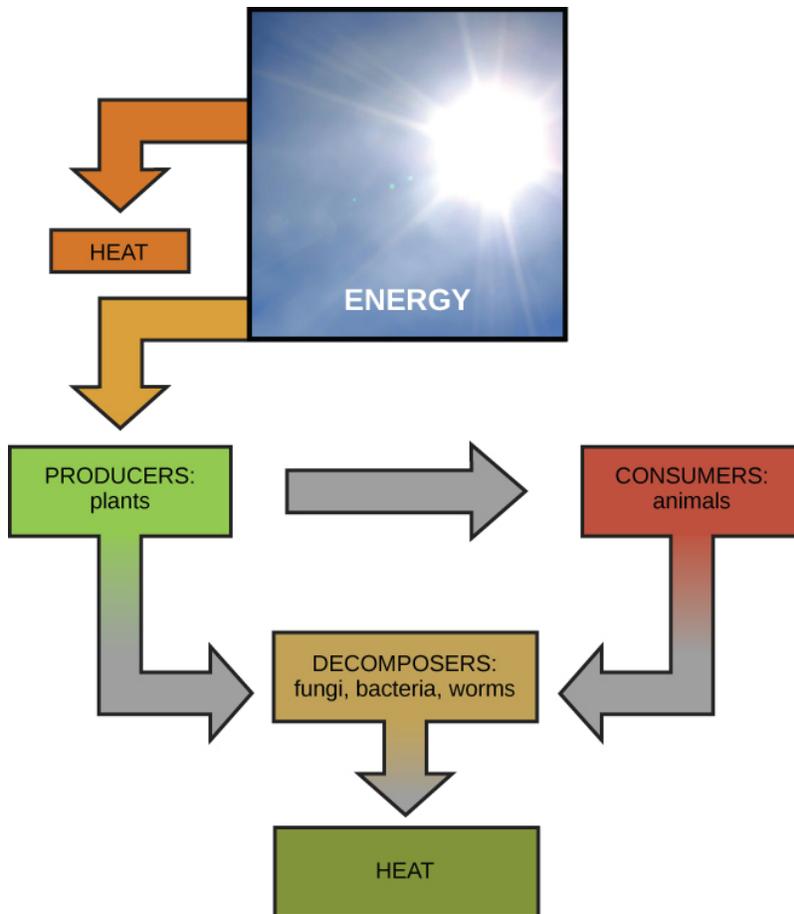
Figure 5.1. Ultimately, most life forms get their energy from the sun. Plants use photosynthesis to capture sunlight, and herbivores eat the plants to obtain energy. Carnivores eat the herbivores, and eventual decomposition of plant and animal material contributes to the nutrient pool.

Metabolic pathways

Consider the metabolism of sugar. This is a classic example of one of the many cellular processes that use and produce energy. Living things consume sugars as a major energy source because sugar molecules have a great deal of energy stored within their bonds. For the most part, photosynthesizing organisms like plants produce these sugars. During photosynthesis, plants use energy (originally from sunlight) to convert carbon dioxide gas (CO_2) into sugar molecules (like glucose: $C_6H_{12}O_6$). They consume carbon dioxide and produce oxygen as a waste product. This reaction is summarized as:

$6CO_2 + 6H_2O$ + energy ---> $C_6H_{12}O_6$+ $6O_2$
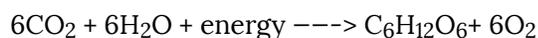
Because this process involves synthesizing an energy-storing molecule, it requires energy input to proceed. During the light reactions of photosynthesis, **energy is provided by a molecule called**

adenosine triphosphate (ATP), which is the primary energy currency of all cells. Just as the dollar is used as currency to buy goods, cells use molecules of ATP as energy currency to perform immediate work. In contrast, energy-storage molecules such as glucose are consumed only to be broken down to use their energy. The reaction that harvests the energy of a sugar molecule in cells requiring oxygen to survive can be summarized by the reverse reaction to photosynthesis. In this reaction, oxygen is consumed and carbon dioxide is released as a waste product. The reaction is summarized as:

$C_6H_{12}O_6 + 6O_2 \longrightarrow 6CO_2 + 6H_2O + energy$

Both of these reactions involve many steps.

The processes of making and breaking down sugar molecules illustrate two examples of metabolic pathways. A metabolic pathway is a series of chemical reactions that takes a starting molecule and modifies it, step-by-step, through a series of metabolic intermediates, eventually yielding a final product. In the example of sugar metabolism, the first metabolic pathway synthesized sugar from smaller molecules, and the other pathway broke sugar down into smaller molecules. These two opposite processes—the first requiring energy and the second producing energy—are referred to as **anabolic pathways (building polymers) and catabolic pathways (breaking down polymers into their monomers)**, respectively. Consequently, metabolism is composed of synthesis (anabolism) and degradation (catabolism) ([Figure 5.2](#)).

It is important to know that the chemical reactions of metabolic pathways do not take place on their own. Each reaction step is facilitated, or catalyzed, by a protein called an enzyme. **Enzymes are important for catalyzing all types of biological reactions**—those that require energy as well as those that release energy.

Metabolic pathways
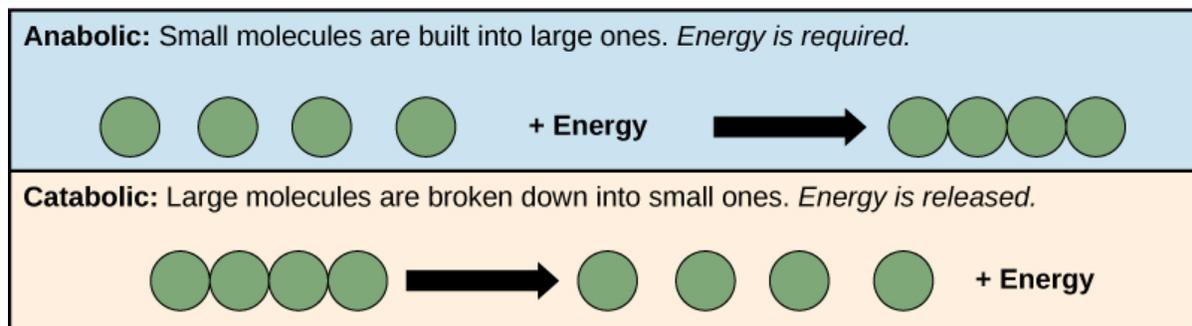
Figure 5.2. Catabolic pathways are those that generate energy by breaking down larger molecules. Anabolic pathways are those that require energy to synthesize larger molecules. Both types of pathways are required for maintaining the cell's energy balance.

Energy

Thermodynamics refers to the study of energy and energy transfer involving physical matter. The matter relevant to a particular case of energy transfer is called a system, and everything outside of that matter is called the surroundings. For instance, when heating a pot of water on the stove, the system includes the stove, the pot, and the water. Energy is transferred within the system (between the stove, pot, and water). There are two types of systems: open and closed. In an open system, energy can be exchanged with its surroundings. The stovetop system is open because heat can be lost to the air. A closed system cannot exchange energy with its surroundings.

Biological organisms are open systems. Energy is exchanged between them and their surroundings as they use energy from the sun to perform photosynthesis or consume energy-storing molecules and release energy to the environment by doing work and releasing heat. Like all things in the physical world, energy is subject to physical laws. The laws of thermodynamics govern the transfer of energy in and among all systems in the universe.

In general, energy is defined as the ability to do work or to create some kind of change. Energy exists in different forms. For example, electrical energy, light energy, and heat energy are all different types of energy. To appreciate the way energy flows into and out of biological systems, it is important to understand two of the physical laws that govern energy.

Thermodynamics

The first law of thermodynamics states that the total amount of energy in the universe is constant and conserved. In other words, there has always been, and always will be, exactly the same amount of energy in the universe. **Energy exists in many different forms**. According to the first law of thermodynamics, energy may be transferred from place to place or transformed into different forms, **but it cannot be created or destroyed**. The transfers and transformations of energy take place around us all the time. Light bulbs transform electrical energy into light and heat energy. Gas stoves transform chemical energy from natural gas into heat energy. Plants perform one of the most biologically useful energy transformations on earth: that of converting the energy of sunlight to chemical energy stored within organic molecules (Figure 5.2). Some examples of energy transformations are shown in [Figure 5.3](#).

The challenge for all living organisms is to obtain energy from their surroundings in forms that they can transfer or transform into usable energy to do work. Living cells have evolved to meet this challenge. Chemical energy stored within organic molecules such as sugars and fats is transferred and transformed through a series of cellular chemical reactions into energy within molecules of ATP. Energy in ATP molecules is easily accessible to do work. Examples of the types of work that cells

need to do include building complex molecules, transporting materials, powering the motion of cilia or flagella, and contracting muscle fibers to create movement.
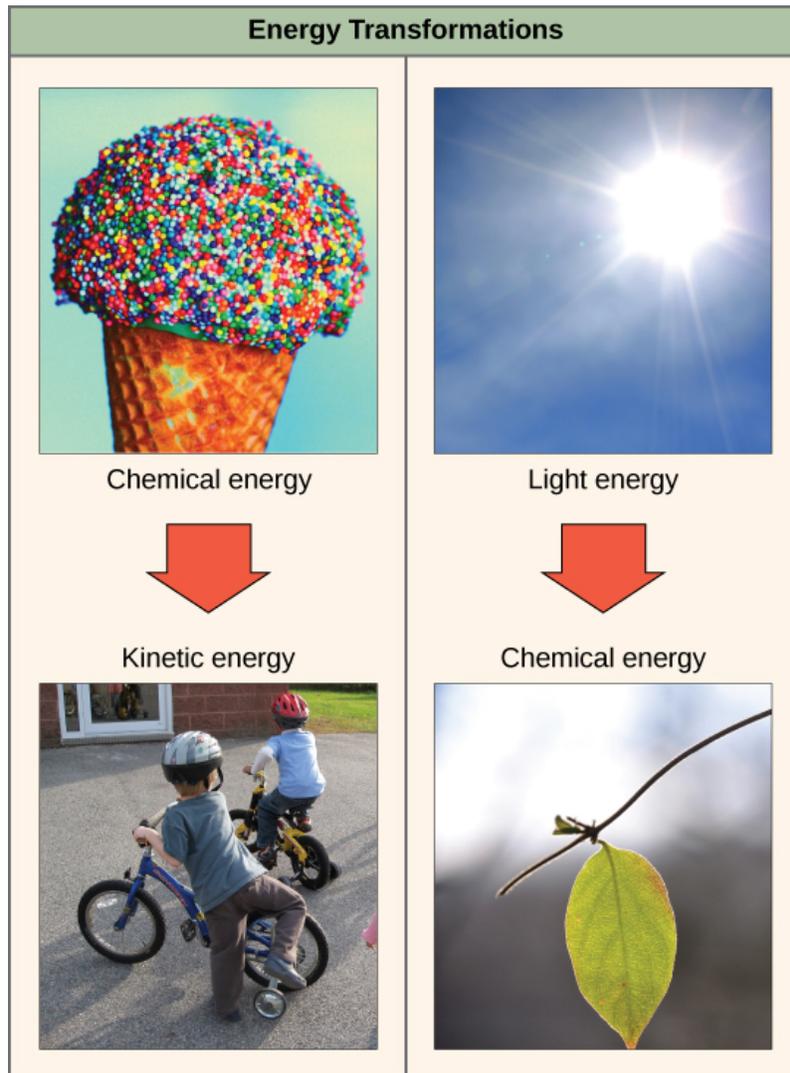

Figure 5.3. Shown are some examples of energy transferred and transformed from one system to another and from one form to another. The food we consume provides our cells with the energy required to carry out bodily functions, just as light energy provides plants with the means to create the chemical energy they need. (credit "ice cream": modification of work by D. Sharon Pruitt; credit "kids": modification of work by Max from Providence; credit "leaf": modification of work by Cory Zanker)

A living cell's primary tasks of obtaining, transforming, and using energy to do work may seem simple. However, the **second law of thermodynamics** explains why these tasks are harder than they appear. **All energy transfers and transformations are never completely efficient**. In every energy transfer, some amount of energy is lost in a form that is unusable. In most cases, this form is heat

energy. Thermodynamically, heat energy is defined as the energy transferred from one system to another that is not work. For example, when a light bulb is turned on, some of the energy being converted from electrical energy into light energy is lost as heat energy. Likewise, some energy is lost as heat energy during cellular metabolic reactions.

An important concept in physical systems is that of order and disorder. The more energy that is lost by a system to its surroundings, the less ordered and more random the system is. Scientists refer to the measure of **randomness or disorder within a system as entropy**. High entropy means high disorder and low energy. Molecules and chemical reactions have varying entropy as well. For example, entropy increases as molecules at a high concentration in one place diffuse and spread out. The second law of thermodynamics says that energy will always be lost as heat in energy transfers or transformations.

Living things are highly ordered, requiring constant energy input to be maintained in a state of low entropy.

Potential and kinetic energy

When an object is in motion, there is energy associated with that object. Think of a wrecking ball. Even a slow-moving wrecking ball can do a great deal of damage to other objects. The energy associated with objects in motion is called kinetic energy (Figure 5.4). A speeding bullet, a walking person, and the rapid movement of molecules in the air (which produces heat) all have kinetic energy.

Now, what if that same motionless wrecking ball is lifted two stories above ground with a crane? If the suspended wrecking ball is unmoving, is there energy associated with it? The answer is yes. The energy that was required to lift the wrecking ball did not disappear but is now stored in the wrecking ball by virtue of its position and the force of gravity acting on it. This type of energy is called potential energy (Figure 5.4). If the ball were to fall, the potential energy would be transformed into kinetic energy until all of the potential energy was exhausted when the ball rested on the ground. Wrecking balls also swing like a pendulum; through the swing, there is a constant change of potential energy (highest at the top of the swing) to kinetic energy (highest at the bottom of the swing). Other examples of potential energy include the energy of water held behind a dam or a person about to skydive out of an airplane.

Figure 5.4. Still water has potential energy; moving water, such as in a waterfall or a rapidly flowing river, has kinetic energy. (credit "dam": modification of work by "Pascal"/Flickr; credit "waterfall": modification of work by Frank Gualtieri).

Potential energy is not only associated with the location of matter, but also with the structure of matter. Even a spring on the ground has potential energy if it is compressed; so does a rubber band that is pulled taut. On a molecular level, the bonds that hold the atoms of molecules together exist in a particular structure that has potential energy. Remember that anabolic cellular pathways require energy to synthesize complex molecules from simpler ones and catabolic pathways release energy when complex molecules are broken down. The fact that energy can be released by the breakdown of certain chemical bonds implies that those bonds have potential energy. In fact, there is potential energy stored within the bonds of all the food molecules we eat, which is eventually harnessed for use. This is because these bonds can release energy when broken. The type of potential energy that exists within chemical bonds, and is released when those bonds are broken, is called chemical energy. Chemical energy is responsible for providing living cells with energy from food. The release of energy occurs when the molecular bonds within food molecules are broken.

Watch this video about kilocalories to help you connect with cellular respiration and energy information you read about in this section of Chapter 5.

Visit this [site](#) and select "Pendulum" from the "Work and Energy" menu to see the shifting kinetic and potential energy of a pendulum in motion.

Free and activation energy

After learning that chemical reactions release energy when energy-storing bonds are broken, an important next question is the following: How is the energy associated with these chemical reactions quantified and expressed? How can the energy released from one reaction be compared to that of another reaction? A measurement of free energy is used to quantify these energy transfers. Recall that according to the second law of thermodynamics, all energy transfers involve the loss of some amount of energy in an unusable form such as heat. Free energy specifically refers to the energy associated with a chemical reaction that is available after the losses are accounted for. In other words, free energy is usable energy or energy that is available to do work.

If energy is released during a chemical reaction, then the change in free energy, signified as ΔG (delta G) will be a negative number. A negative change in free energy also means that the products of the reaction have less free energy than the reactants because they release some free energy during the reaction. Reactions that have a negative change in free energy and consequently release free energy are called exergonic reactions. Think: **exergonic** means energy is *ex*iting the system. These reactions are also referred to as spontaneous reactions, and their products have less stored energy than the reactants. An important distinction must be drawn between the term spontaneous and the idea of a chemical reaction occurring immediately. Contrary to the everyday use of the term, a spontaneous reaction is not one that suddenly or quickly occurs. The rusting of iron is an example of a spontaneous reaction that occurs slowly, little by little, over time.

If a chemical reaction absorbs energy rather than releases energy on balance, then the ΔG for that reaction will be a positive value. In this case, the products have more free energy than the reactants. Thus, the products of these reactions can be thought of as energy-storing molecules. These chemical reactions are called **endergonic** reactions and they are **non-spontaneous**. An endergonic reaction will not take place on its own without the addition of free energy.

Look at each of the processes shown in Figure 5.5 and decide if it is endergonic or exergonic.

(a)

(b)

(c)

(d)

Figure 5.5. Some examples of endergonic processes (ones that require energy) and exergonic processes (ones that release energy). (credit a: modification of work by Natalie Maynor; credit b: modification of work by USDA; credit c: modification of work by Cory Zanker; credit d: modification of work by Harry Malsch).

There is another important concept that must be considered regarding endergonic and exergonic reactions. Exergonic reactions require a small amount of energy input to get going before they can proceed with their energy-releasing steps. These reactions have a net release of energy, but still,

require some energy input in the beginning. This small amount of energy input necessary for all chemical reactions to occur is called the activation energy.

Enzymes

A substance that helps a chemical reaction to occur is called a catalyst, and the molecules that catalyze biochemical reactions are called enzymes. Most enzymes are **proteins** and perform the critical task of **lowering the activation energies** of chemical reactions inside the cell. Most of the reactions critical to a living cell happen too slowly at normal temperatures to be of any use to the cell. Without enzymes to **speed up these reactions**, life could not persist. Enzymes do this by binding to the reactant molecules and holding them in such a way as to make the chemical bond-breaking and -forming processes take place more easily. It is important to remember that enzymes do not change whether a reaction is exergonic (spontaneous) or endergonic. This is because they do not change the free energy of the reactants or products. They only reduce the activation energy required for the reaction to go forward (Figure 5.6). In addition, an enzyme itself is unchanged by the reaction it catalyzes. Once one reaction has been catalyzed, the enzyme is able to participate in other reactions.
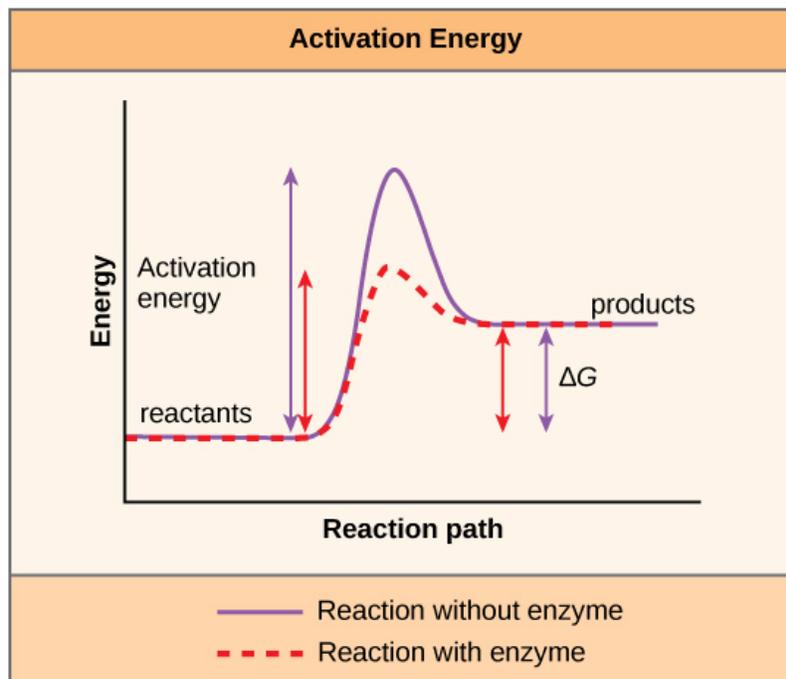
Figure 5.6. Enzymes lower the activation energy of the reaction but do not change the free energy of the reaction.

The chemical reactants to which an enzyme binds are called the enzyme's substrates. There may be one or more substrates, depending on the particular chemical reaction. In some reactions, a

single reactant substrate is broken down into multiple products. In others, two substrates may come together to create one larger molecule. Two reactants might also enter a reaction and both become modified, but they leave the reaction as two products. The location within the enzyme where the substrate binds is called the enzyme's **active site**. The active site is where the "action" happens. Since enzymes are proteins, there is a unique combination of amino acid side chains within the active site. Each side chain is characterized by different properties. They can be large or small, weakly acidic or basic, hydrophilic or hydrophobic, positively or negatively charged, or neutral. The unique combination of side chains creates a very specific chemical environment within the active site. This specific environment is suited to bind to one specific chemical substrate (or substrates).

Active sites are subject to influences of the local environment. Increasing the environmental temperature generally increases reaction rates, enzyme-catalyzed or otherwise. However, temperatures outside of an optimal range reduce the rate at which an enzyme catalyzes a reaction. Hot temperatures will eventually cause enzymes to denature, an irreversible change in the three-dimensional shape and therefore the function of the enzyme. Enzymes are also suited to function best within a certain pH and salt concentration range, and, as with temperature, extreme pH, and salt concentrations can cause enzymes to denature.

For many years, scientists thought that enzyme-substrate binding took place in a simple "lock and key" fashion. This model asserted that the enzyme and substrate fit together perfectly in one instantaneous step. However, current research supports a model called induced fit (Figure 5.7). The induced-fit model expands on the lock-and-key model by describing a more dynamic binding between enzyme and substrate. As the enzyme and substrate come together, their interaction causes a mild shift in the enzyme's structure that forms an ideal binding arrangement between enzyme and substrate.

View an animation of induced fit.

When an enzyme binds its substrate, an enzyme-substrate complex is formed. This complex lowers the activation energy of the reaction and promotes its rapid progression in one of the multiple possible ways. On a basic level, enzymes promote chemical reactions that involve more than one substrate by bringing the substrates together in an optimal orientation for reaction. Another way in which enzymes promote the reaction of their substrates is by creating an optimal environment within the active site for the reaction to occur. The chemical properties that emerge from the particular arrangement of amino acid R groups within an active site create the perfect environment for an enzyme's specific substrates to react.

The enzyme-substrate complex can also lower activation energy by compromising the bond structure so that it is easier to break. Finally, enzymes can also lower activation energies by taking part in the chemical reaction itself. In these cases, it is important to remember that the enzyme will always return to its original state upon the completion of the reaction. One of the hallmark properties of enzymes is that they remain ultimately unchanged by the reactions they catalyze. After an enzyme has catalyzed a reaction, it releases its product(s) and can catalyze a new reaction.
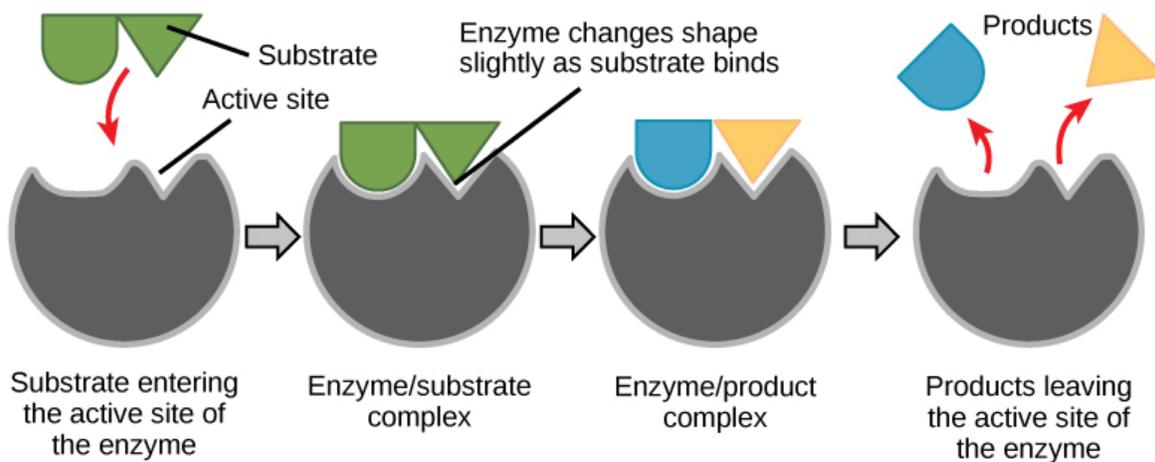
Figure 5.7. The induced-fit model is an adjustment to the lock-and-key model and explains how enzymes and substrates undergo dynamic modifications during the transition state to increase the affinity of the substrate for the active site.

It would seem ideal to have a scenario in which all of an organism's enzymes existed in abundant supply and functioned optimally under all cellular conditions, in all cells, at all times. However, a variety of mechanisms ensures that this does not happen. Cellular needs and conditions constantly vary from cell to cell and change within individual cells over time. The required enzymes of stomach cells differ from those of fat storage cells, skin cells, blood cells, and nerve cells. Furthermore, a digestive organ cell works much harder to process and break down nutrients during the time that closely follows a meal compared with many hours after a meal. As these cellular demands and conditions vary, so must the amounts and functionality of different enzymes.

Since the rates of biochemical reactions are controlled by activation energy, and enzymes lower and determine activation energies for chemical reactions, the relative amounts and functioning of the variety of enzymes within a cell ultimately determine which reactions will proceed and at what rates. This determination is tightly controlled in cells. In certain cellular environments, enzyme activity is partly controlled by environmental factors like pH, temperature, salt concentration, and, in some cases, cofactors or coenzymes.

Enzymes can also be regulated in ways that either promote or reduce enzyme activity. There are many kinds of molecules that inhibit or promote enzyme function, and various mechanisms by which they do so. In some cases of enzyme **inhibition**, an inhibitor molecule is similar enough to a substrate

that it can bind to the active site and simply block the substrate from binding. When this happens, the enzyme is inhibited through **competitive inhibition**, because an inhibitor molecule competes with the substrate for binding to the active site.

On the other hand, in **noncompetitive inhibition**, an inhibitor molecule binds to the enzyme in a location other than the active site, called an **allosteric site**, but still manages to block substrate binding to the active site. Some inhibitor molecules bind to enzymes in a location where their binding induces a conformational change that reduces the affinity of the enzyme for its substrate. This type of inhibition is called allosteric inhibition (Figure 5.8). Most allosterically regulated enzymes are made up of more than one polypeptide, meaning that they have more than one protein subunit. When an allosteric inhibitor binds to a region on an enzyme, all active sites on the protein subunits are changed slightly such that they bind their substrates with less efficiency. There are allosteric activators as well as inhibitors. Allosteric activators bind to locations on an enzyme away from the active site, inducing a conformational change that increases the affinity of the enzyme's active site(s) for its substrate(s) ([Figure 5.8](#)).
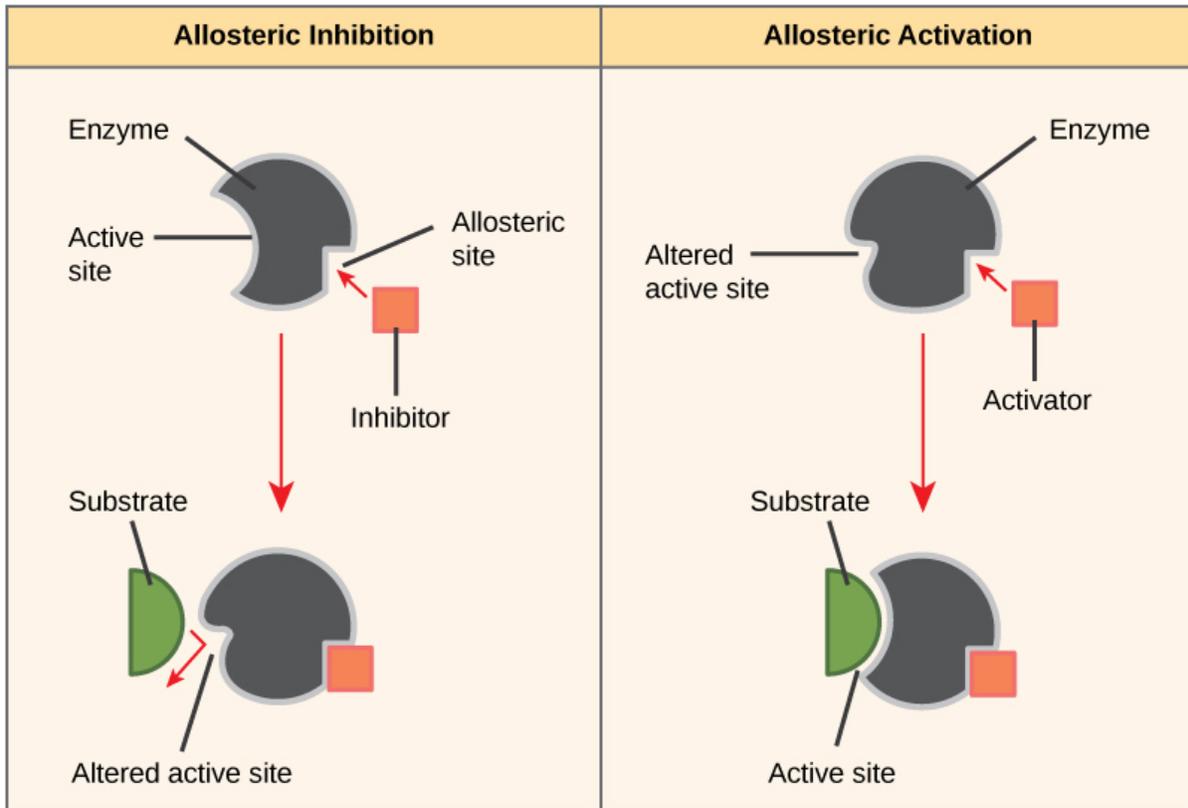
Figure 5.8. Allosteric inhibition works by indirectly inducing a conformational change to the active site such that the substrate no longer fits. In contrast, in allosteric activation, the activator molecule modifies the shape of the active site to allow a better fit of the substrate.

Enzymes are key components of metabolic pathways. Understanding how enzymes work and how

they can be regulated are key principles behind the development of many of the pharmaceutical drugs on the market today. Biologists working in this field collaborate with other scientists to design drugs ([Figure 5.9](#)).

Figure 5.9. Pharmaceutical drug developer. Have you ever wondered how pharmaceutical drugs are developed? (credit: Deborah Austin)

Consider statins for example–statins is the name given to one class of drugs that can reduce cholesterol levels. These compounds are inhibitors of the enzyme HMG-CoA reductase, which is the enzyme that synthesizes cholesterol from lipids in the body. By inhibiting this enzyme, the level of cholesterol synthesized in the body can be reduced. Similarly, acetaminophen, popularly marketed under the brand name Tylenol, is an inhibitor of the enzyme cyclooxygenase. While it is used to provide relief from fever and inflammation (pain), its mechanism of action is still not completely understood.

How are drugs discovered? One of the biggest challenges in drug discovery is identifying a drug target. A drug target is a molecule that is literally the target of the drug. In the case of statins, HMG-CoA reductase is the drug target. Drug targets are identified through painstaking research in the laboratory. Identifying the target alone is not enough; scientists also need to know how the target acts inside the cell and which reactions go awry in the case of a disease. Once the target and the pathway are identified, then the actual process of drug design begins. In this stage, chemists and biologists work together to design and synthesize molecules that can block or activate a particular reaction. However, this is only the beginning: If and when a drug prototype is successful in performing its function, then it is subjected to many tests from in vitro experiments to clinical trials before it can get approval from the U.S. Food and Drug Administration to be on the market.

Many enzymes do not work optimally, or even at all, unless bound to other specific non-protein helper molecules. They may bond either temporarily through ionic or hydrogen bonds, or permanently through stronger covalent bonds. Binding to these molecules promotes optimal shape and function of their respective enzymes. Two examples of these types of helper molecules are cofactors and coenzymes. Cofactors are inorganic ions such as ions of iron and magnesium. Coenzymes are organic helper molecules, those with a basic atomic structure made up of carbon and hydrogen. Like enzymes, these molecules participate in reactions without being changed themselves and are ultimately recycled and reused. Vitamins are the source of coenzymes. Some vitamins are the precursors of coenzymes and others act directly as coenzymes. Vitamin C is a direct coenzyme for multiple enzymes that take part in building the important connective tissue, collagen. Therefore, enzyme function is, in part, regulated by the abundance of various cofactors and coenzymes, which may be supplied by an organism's diet or, in some cases, produced by the organism.

Feedback inhibition in metabolic pathways

Molecules can regulate enzyme function in many ways. The major question remains, however: What are these molecules and where do they come from? Some are cofactors and coenzymes, as you have learned. What other molecules in the cell provide enzymatic regulation such as allosteric modulation and competitive and non-competitive inhibition? Perhaps the most relevant sources of regulatory molecules, with respect to enzymatic cellular metabolism, are the products of the cellular metabolic reactions themselves. In a most efficient and elegant way, cells have evolved to use the products of their own reactions for feedback inhibition of enzyme activity. Feedback inhibition involves the use of a reaction product to regulate its own further production (Figure 5.10). The cell responds to an abundance of the products by slowing down production during anabolic or catabolic reactions. Such reaction products may inhibit the enzymes that catalyzed their production through the mechanisms described above.
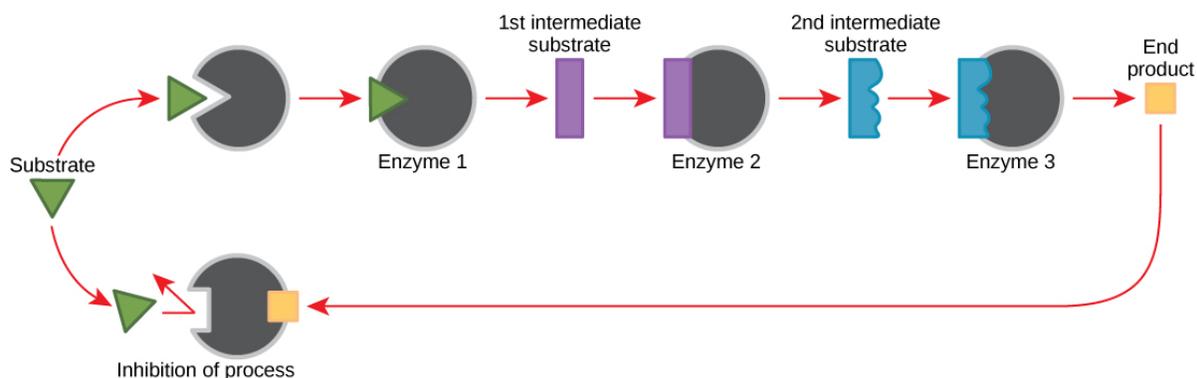

Figure 5.10. Metabolic pathways are a series of reactions catalyzed by multiple enzymes. Feedback inhibition, where the end product of the pathway inhibits an upstream process, is an important regulatory mechanism in cells.

The production of both amino acids and nucleotides is controlled through feedback inhibition. Additionally, ATP is an allosteric regulator of some of the enzymes involved in the catabolic breakdown of sugar, the process that creates ATP. In this way, when ATP is in abundant supply, the cell can prevent the production of ATP. On the other hand, ADP serves as a positive allosteric regulator (an allosteric activator) for some of the same enzymes that are inhibited by ATP. Thus, when relative levels of ADP are high compared to ATP, the cell is triggered to produce more ATP through sugar catabolism.

Question 5.2

What are different types of energy and which one is most commonly used by animals to sustain and maintain life?

Question 5.3

How would energy production and needs be affected by body size of an animal? What about its environment (aquatic versus terrestrial)?

5.2 – Components of the Digestive System

5.2. Compare and contrast digestive system structures in sample animal systems.

Animals maintain work and energy demands on their body through nutrition. They obtain their nutrition from the consumption of other organisms. Depending on their diet, animals can be classified into the following categories: plant eaters (herbivores), meat eaters (carnivores), and those that eat both plants and animals (omnivores). The nutrients and macromolecules present in food are not immediately accessible to the cells. There are a number of processes that modify food within the animal body in order to make the nutrients and organic molecules accessible for cellular function. As animals evolved in the complexity of form and function, their digestive systems have also evolved to accommodate their various dietary needs.

Herbivores, omnivores, and carnivores

Herbivores are animals whose primary food source is plant-based. Examples of herbivores, as shown in Figure 5.11 include vertebrates like deer, koalas, and some bird species, as well as invertebrates such as crickets and caterpillars. These animals have evolved digestive systems capable of handling large amounts of plant material. Herbivores can be further classified into frugivores (fruit-eaters), granivores (seed eaters), nectivores (nectar feeders), and folivores (leaf eaters).

Figure 5.11. Herbivores, like this (a) mule deer and (b) monarch caterpillar, eat primarily plant material. (credit a: modification of work by Bill Ebbesen; credit b: modification of work by Doug Bowman)

Carnivores are animals that eat other animals. The word carnivore is derived from Latin and literally means "meat eater." Wild cats such as lions, shown in Figure 5.12a and tigers are examples of vertebrate carnivores, as are snakes and sharks, while invertebrate carnivores include sea stars, spiders, and ladybugs, shown in Figure 5.12b. Obligate carnivores are those that rely entirely on animal flesh to obtain their nutrients; examples of obligate carnivores are members of the cat family, such as lions and cheetahs. Facultative carnivores are those that also eat non-animal food in addition to animal food. Note that there is no clear line that differentiates facultative carnivores from omnivores; dogs would be considered facultative carnivores.

(a) (b)

Figure 5.12. Carnivores like the (a) lion eat primarily meat. The (b) ladybug is also a carnivore that consumes small insects called aphids. (credit a: modification of work by Kevin Pluck; credit b: modification of work by Jon Sullivan)

Omnivores are animals that eat both plant- and animal-derived food. In Latin, omnivore means to eat everything. Humans, bears (shown in Figure 5.13a), and chickens are examples of vertebrate omnivores; invertebrate omnivores include cockroaches and crayfish (shown in Figure 5.13b).
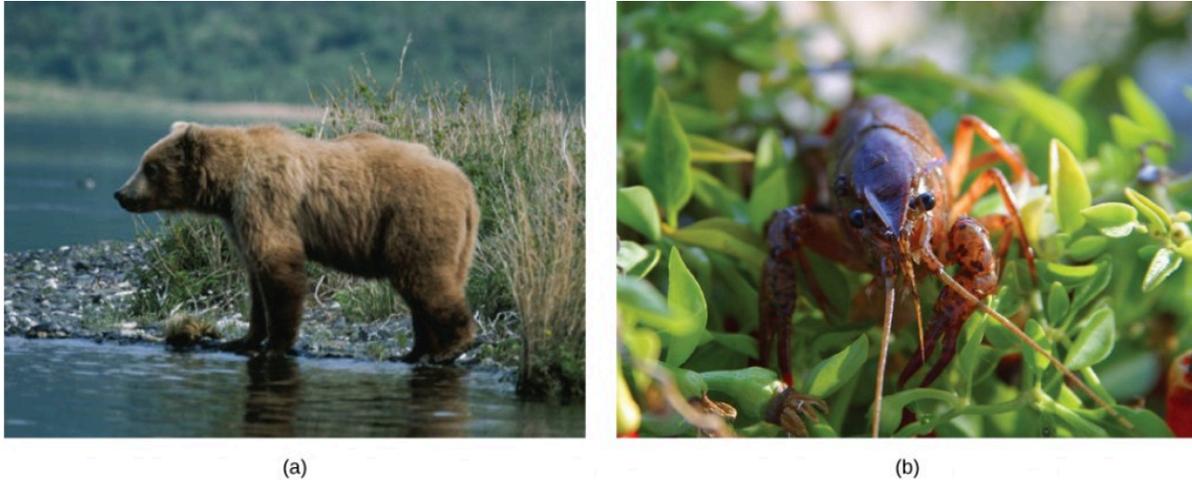
Figure 5.13. Omnivores like the (a) bear and (b) crayfish eat both plant and animal-based food. (credit a: modification of work by Dave Menke; credit b: modification of work by Jon Sullivan)

Invertebrate digestive systems

Animals have evolved different types of digestive systems to aid in the digestion of the different foods they consume. The simplest example is that of a **gastrovascular cavity** and is found in organisms with only one opening for digestion. Platyhelminthes (flatworms), Ctenophora (comb jellies), and Cnidaria (coral, jellyfish, and sea anemones) use this type of digestion. Gastrovascular cavities, as shown in Figure 5.14a, are typically a blind tube or cavity with only one opening, the "mouth", which also serves as an "anus". Ingested material enters the mouth and passes through a hollow, tubular cavity. Cells within the cavity secrete digestive enzymes that break down the food. The food particles are engulfed by the cells lining the gastrovascular cavity.

The **alimentary canal**, shown in Figure 5.14b, is a more advanced system: it consists of one tube with a mouth at one end and an anus at the other. Earthworms are an example of an animal with an alimentary canal. Once the food is ingested through the mouth, it passes through the esophagus and is stored in an organ called the crop; then it passes into the gizzard where it is churned and digested. From the gizzard, the food passes through the intestine, the nutrients are absorbed, and the waste is eliminated as feces, called castings, through the anus.
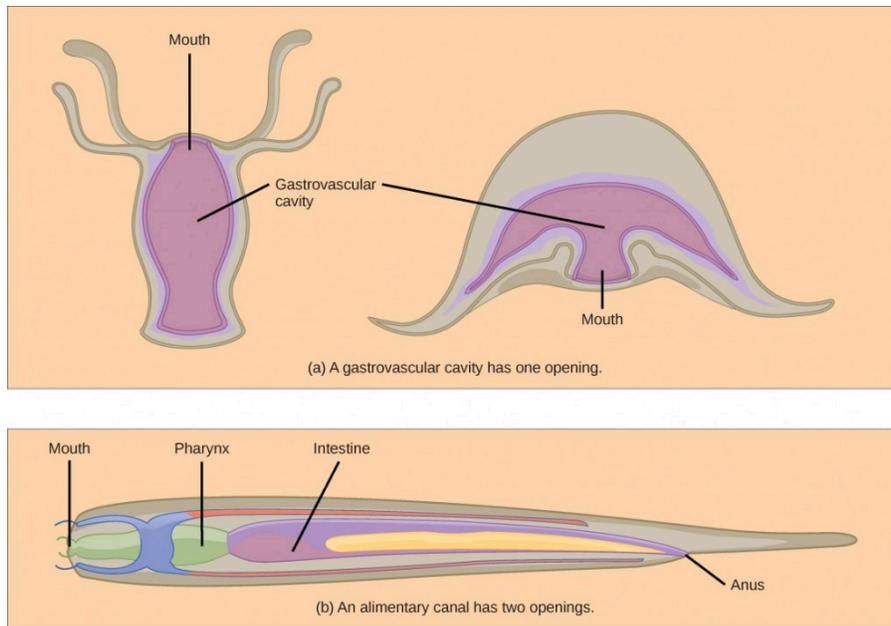
Figure 5.14. (a) A gastrovascular cavity has a single opening through which food is ingested and waste is excreted, as shown in this hydra and in this jellyfish medusa. (b) An alimentary canal has two openings: a mouth for ingesting food, and an anus for eliminating waste, as shown in this nematode.

Vertebrate digestive systems

Vertebrates have evolved more complex digestive systems to adapt to their dietary needs. Some animals have a single stomach, while others have multi-chambered stomachs. Birds have developed a digestive system adapted to eating unmasticated food.

Monogastric: single-chambered stomach

As the word **monogastric** suggests, this type of digestive system consists of one ("mono") stomach chamber ("gastric"). Humans and many animals have a monogastric digestive system as illustrated in Figure 5.15. The process of digestion begins with the mouth and the intake of food. The teeth play an important role in masticating (chewing) or physically breaking down food into smaller particles. The enzymes present in saliva also begin to chemically break down food. The esophagus is a long tube that connects the mouth to the stomach. Using peristalsis, or wave-like smooth muscle contractions, the muscles of the esophagus push the food towards the stomach. In order to speed up the actions of enzymes in the stomach, the stomach is an extremely acidic environment, with a pH between 1.5 and 2.5. The gastric juices, which include enzymes in the stomach, act on the food particles and

continue the process of digestion. The further breakdown of food takes place in the small intestine where enzymes produced by the liver, the small intestine, and the pancreas continue the process of digestion. The nutrients are absorbed into the bloodstream across the epithelial cells lining the walls of the small intestines. The waste material travels on to the large intestine where water is absorbed and the drier waste material is compacted into feces; it is stored until it is excreted through the rectum.
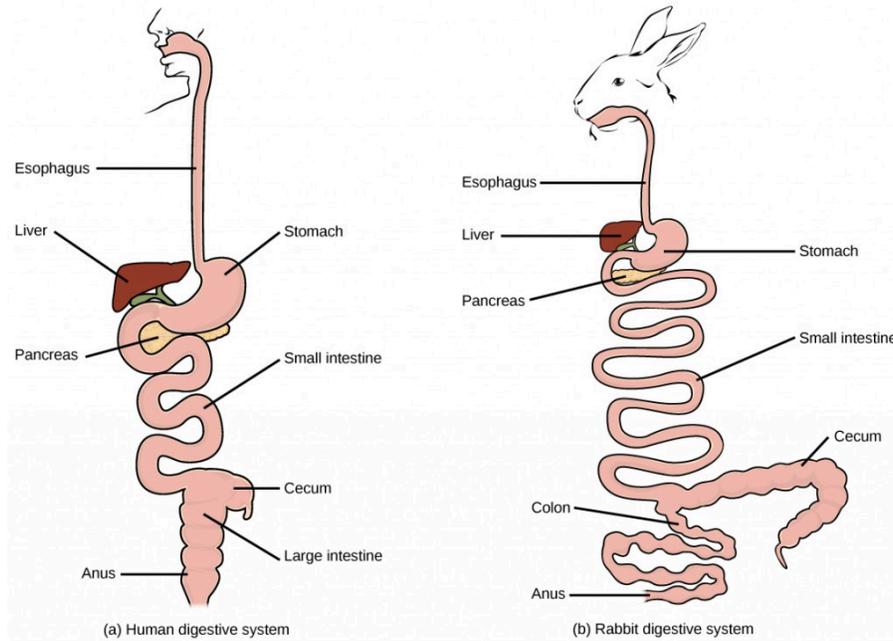

Figure 5.15. (a) Humans and herbivores, such as the (b) rabbit, have a monogastric digestive system. However, in the rabbit, the small intestine and cecum are enlarged to allow more time to digest plant material. The enlarged organ provides more surface area for absorption of nutrients. Rabbits digest their food twice: the first time food passes through the digestive system, it collects in the cecum, and then it passes as soft feces called cecotrophes. The rabbit re-ingests these cecotrophes to further digest them.

Avian digestive system

Birds face special challenges when it comes to obtaining nutrition from food. They do not have teeth and so their digestive system, shown in Figure 5.16, must be able to process un-masticated food. Birds have evolved a variety of beak types that reflect the vast variety in their diet, ranging from seeds and insects to fruits and nuts. Because most birds fly, their metabolic rates are high in order to efficiently process food and keep their body weight low. The stomach of birds has two chambers: the **proventriculus**, where gastric juices are produced to digest the food before it enters the stomach, and the **gizzard**, where the food is stored, soaked, and mechanically ground. The undigested material

forms food pellets that are sometimes regurgitated. Most of the chemical digestion and absorption happens in the intestine and the waste is excreted through the cloaca.
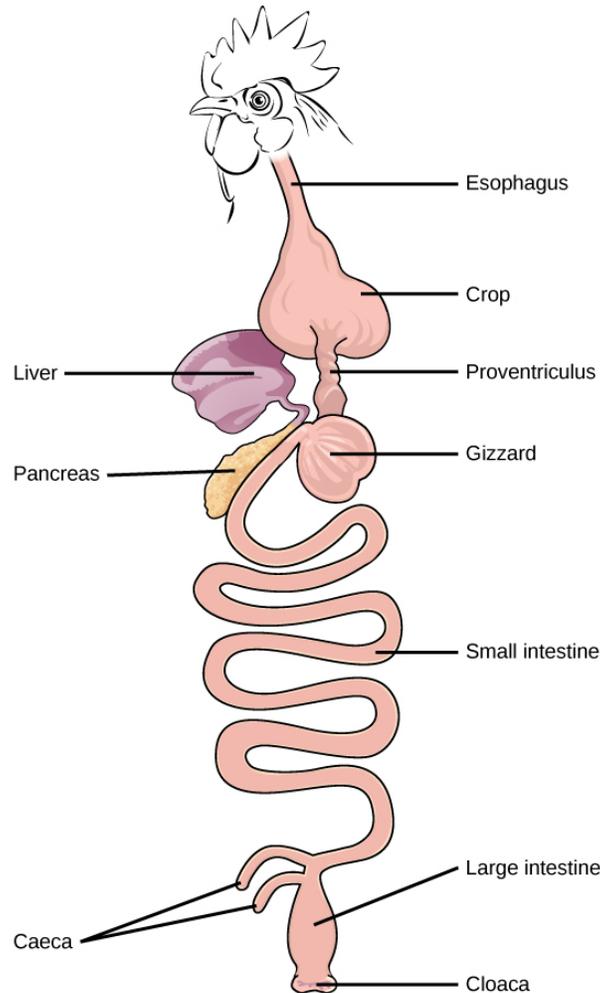

Figure 5.16. The avian esophagus has a pouch, called a crop, which stores food. Food passes from the crop to the first of two stomachs, called the proventriculus, which contains digestive juices that break down food. From the proventriculus, the food enters the second stomach, called the gizzard, which grinds food. Some birds swallow stones or grit, which are stored in the gizzard, to aid the grinding process. Birds do not have separate openings to excrete urine and feces. Instead, uric acid from the kidneys is secreted into the large intestine and combined with waste from the digestive process. This waste is excreted through an opening called the cloaca.

Parts of the digestive system

The vertebrate digestive system is designed to facilitate the transformation of food matter into the nutrient components that sustain organisms.

Oral cavity

The oral cavity, or mouth, is the point of entry of food into the digestive system, illustrated in [Figure 5.17](#). The food consumed is broken into smaller particles by mastication, the chewing action of the teeth. All mammals have teeth and can chew their food

The extensive chemical process of digestion begins in the mouth. As food is being chewed, saliva, produced by the salivary glands, mixes with the food. Saliva is a watery substance produced in the mouths of many animals. There are three major glands that secrete saliva—the parotid, the submandibular, and the sublingual. Saliva contains mucus that moistens food and buffers the pH of the food. Saliva also contains immunoglobulins and lysozymes, which have an antibacterial action to reduce tooth decay by inhibiting the growth of some bacteria. Saliva also contains an enzyme called **salivary amylase** that begins the process of converting starches in the food into a disaccharide called maltose. Another enzyme called **lipase** is produced by the cells in the tongue. Lipases are a class of enzymes that can break down triglycerides. The lingual lipase begins the breakdown of fat components in the food. The chewing and wetting action provided by the teeth and saliva prepare the food into a mass called the **bolus** for swallowing. The tongue helps in swallowing—moving the bolus from the mouth into the pharynx. The pharynx opens to two passageways: the trachea, which leads to the lungs, and the esophagus, which leads to the stomach. The trachea has an opening called the glottis, which is covered by a cartilaginous flap called the epiglottis. When swallowing, the epiglottis closes the glottis and food passes into the esophagus and not the trachea. This arrangement allows food to be kept out of the trachea.
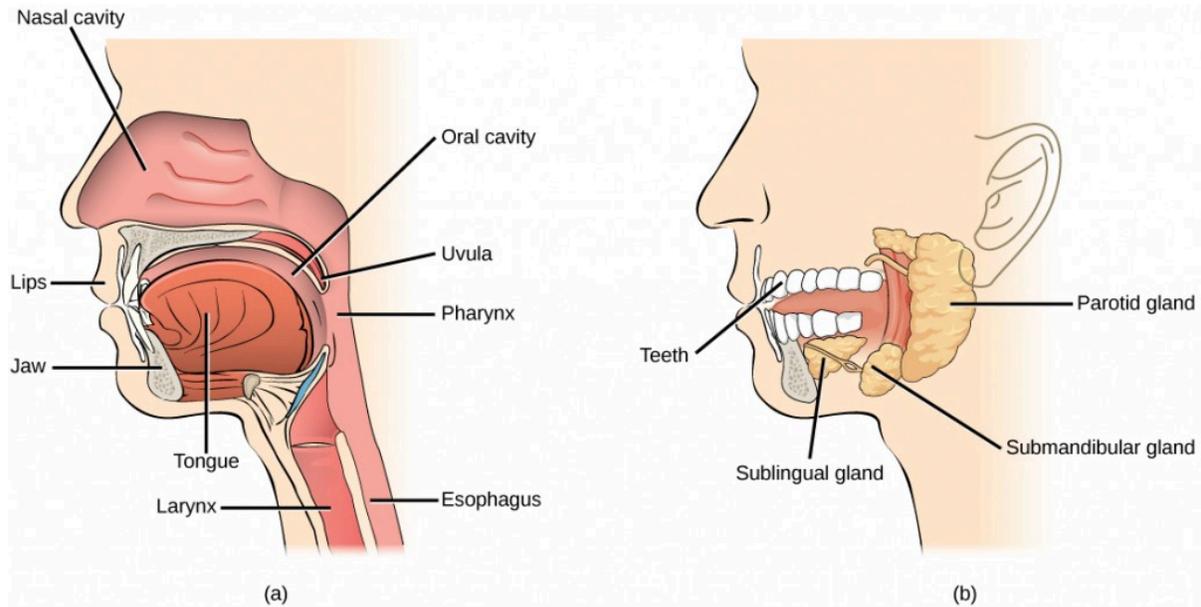
Figure 5.17. Digestion of food begins in the (a) oral cavity. Food is masticated by teeth and moistened by saliva secreted from the (b) salivary glands. Enzymes in the saliva begin to digest starches and fats. With the help of the tongue, the resulting bolus is moved into the esophagus by swallowing. (credit: modification of work by the National Cancer Institute)

Esophagus

The **esophagus** is a tubular organ that connects the mouth to the stomach. The chewed and softened food passes through the esophagus after being swallowed. The smooth muscles of the esophagus undergo a series of wave-like movements called **peristalsis** that push the food toward the stomach, as illustrated in Figure 5.18. The peristalsis wave is unidirectional—it moves food from the mouth to the stomach, and reverse movement is not possible. The peristaltic movement of the esophagus is an involuntary reflex; it takes place in response to the act of swallowing.
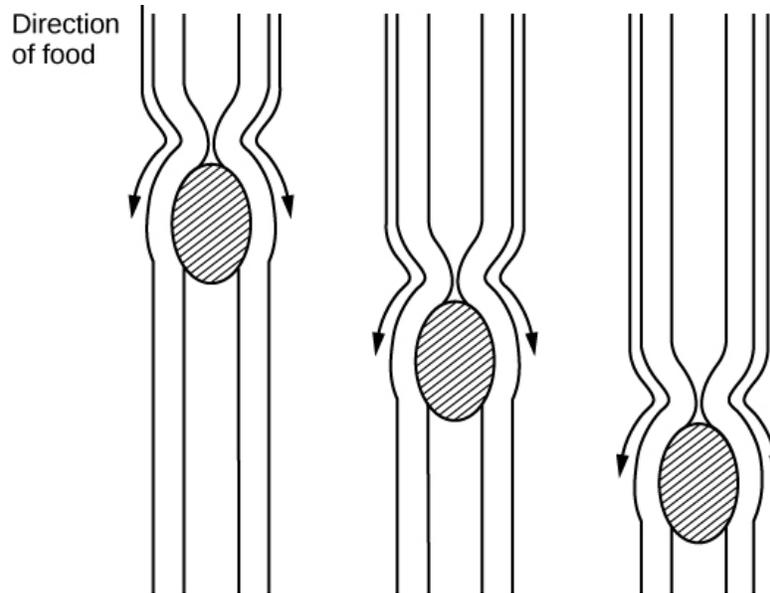
Figure 5.18. The esophagus transfers food from the mouth to the stomach through peristaltic movements.

A ring-like muscle called a **sphincter** forms valves in the digestive system. The gastro-esophageal sphincter is located at the stomach end of the esophagus. In response to swallowing and the pressure exerted by the bolus of food, this sphincter opens, and the bolus enters the stomach. When there is no swallowing action, this sphincter is shut and prevents the contents of the stomach from traveling up the esophagus. Many animals have a true sphincter; however, in humans, there is no true sphincter, but the esophagus remains closed when there is no swallowing action. Acid reflux or "heartburn" occurs when the acidic digestive juices escape into the esophagus.

Stomach

A large part of digestion occurs in the stomach, shown in Figure 5.19. The **stomach** is a sac-like organ that secretes gastric digestive juices. The pH in the stomach is between 1.5 and 2.5. This highly acidic environment is required for the chemical breakdown of food and the extraction of nutrients. When empty, the stomach is a rather small organ; however, it can expand to up to 20 times its resting size when filled with food. This characteristic is particularly useful for animals that need to eat when food is available.
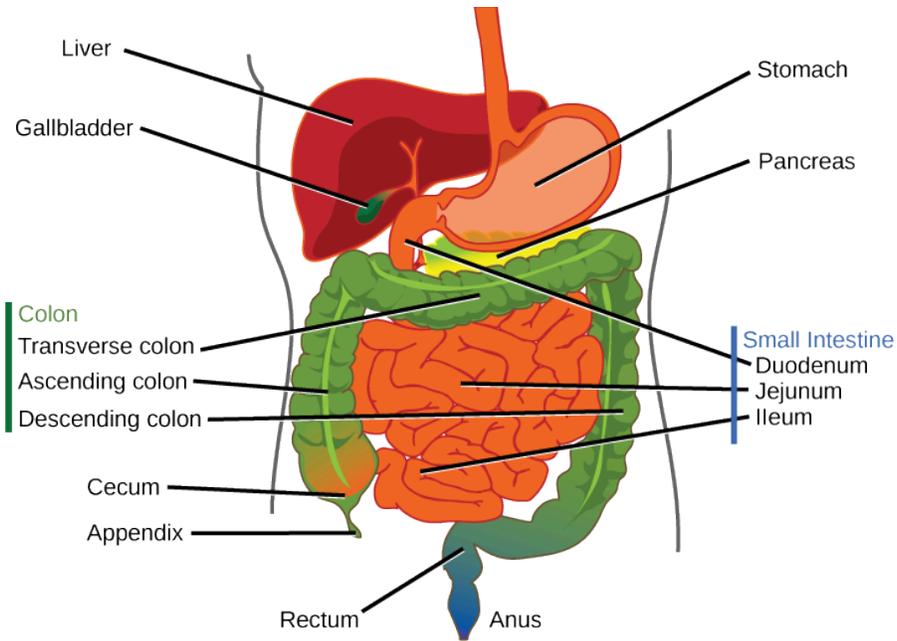
Figure 5.19. The human stomach has an extremely acidic environment where most of the protein gets digested. (credit: modification of work by Mariana Ruiz Villareal)

Question 5.4

Which of the following statements about the digestive system is false?
a. Chyme is a mixture of food and digestive juices that is produced in the stomach.
b. Food enters the large intestine before the small intestine.
c. In the small intestine, chyme mixes with bile, which emulsifies fats.
d. The stomach is separated from the small intestine by the pyloric sphincter.

The stomach is also the major site for protein digestion in animals other than ruminants. Protein digestion is mediated by an enzyme called pepsin in the stomach chamber. **Pepsin** is secreted by the chief cells in the stomach in an inactive form called **pepsinogen**. Pepsin breaks peptide bonds and cleaves proteins into smaller polypeptides; it also helps activate more pepsinogen, starting a positive feedback mechanism that generates more pepsin. Another cell type–parietal cells–secrete hydrogen and chloride ions, which combine in the lumen to form hydrochloric acid, the primary acidic component of the stomach juices. Hydrochloric acid helps to convert the inactive pepsinogen to pepsin. The highly acidic environment also kills many microorganisms in the food and, combined with the action of the enzyme pepsin, results in the hydrolysis of protein in the food. Chemical digestion is facilitated by the churning action of the stomach. Contraction and relaxation of smooth muscles mix the stomach contents about every 20 minutes. The partially digested food and gastric

juice mixture are called **chyme**. Chyme passes from the stomach to the small intestine. Further protein digestion takes place in the small intestine. Gastric emptying occurs within two to six hours after a meal. Only a small amount of chyme is released into the small intestine at a time. The movement of chyme from the stomach into the small intestine is regulated by the pyloric sphincter.

When digesting protein and some fats, the stomach lining must be protected from getting digested by pepsin. There are two points to consider when describing how the stomach lining is protected. First, as previously mentioned, the enzyme pepsin is synthesized in the inactive form. This protects the chief cells because pepsinogen does not have the same enzyme functionality of pepsin. Second, the stomach has a thick mucus lining that protects the underlying tissue from the action of the digestive juices. When this mucus lining is ruptured, ulcers can form in the stomach. Ulcers are open wounds in or on an organ caused by bacteria (*Helicobacter pylori*) when the mucus lining is ruptured and fails to reform.

Small intestine

Chyme moves from the stomach to the small intestine. The **small intestine** is the organ where the digestion of protein, fats, and carbohydrates is completed. The small intestine is a long tube-like organ with a highly folded surface containing finger-like projections called the **villi**. The apical surface of each villus has many microscopic projections called microvilli. These structures, illustrated in Figure 5.20, are lined with epithelial cells on the luminal side and allow for the nutrients to be absorbed from the digested food and absorbed into the blood stream on the other side. The villi and microvilli, with their many folds, increase the surface area of the intestine and increase absorption efficiency of the nutrients. Absorbed nutrients in the blood are carried into the hepatic portal vein, which leads to the liver. There, the liver regulates the distribution of nutrients to the rest of the body and removes toxic substances, including drugs, alcohol, and some pathogens.
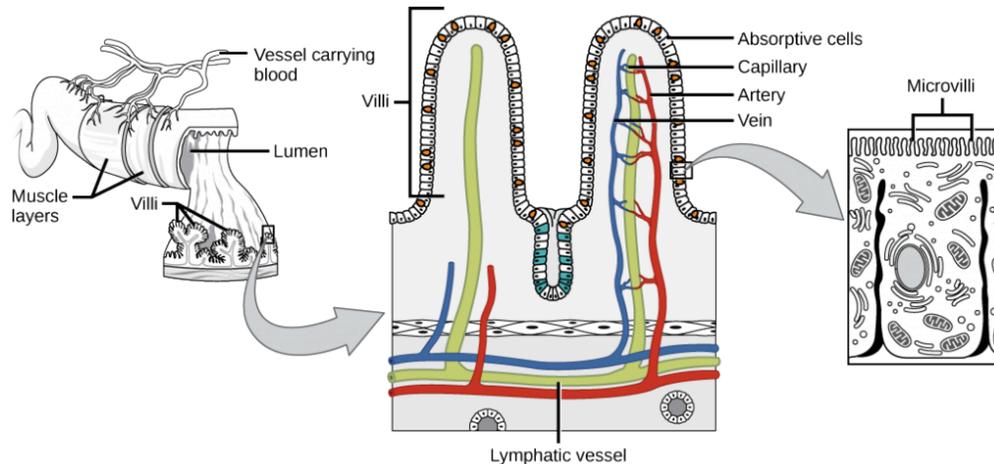
Figure 5.20. Villi are folds on the small intestine lining that increase the surface area to facilitate the absorption of nutrients.

Question 5.5

Which of the following statements about the small intestine is false?
a. Absorptive cells that line the small intestine have microvilli, small projections that increase surface area and aid in the absorption of food.
b. The inside of the small intestine has many folds, called villi.
c. Microvilli are lined with blood vessels as well as lymphatic vessels.
d. The inside of the small intestine is called the lumen.

The human small intestine is over 6m long and is divided into three parts: the duodenum, the jejunum, and the ileum. The "C-shaped," fixed part of the small intestine is called the **duodenum** and is shown in Figure 5.19. The duodenum is separated from the stomach by the pyloric sphincter which opens to allow chyme to move from the stomach to the duodenum. In the duodenum, chyme is mixed with pancreatic juices in an alkaline solution rich in bicarbonate that neutralizes the acidity of chyme and acts as a buffer. Pancreatic juices also contain several digestive enzymes. Digestive juices from the pancreas, liver, and gallbladder, as well as from gland cells of the intestinal wall itself, enter the duodenum. **Bile** is produced in the liver and stored and concentrated in the gallbladder. Bile contains bile salts which emulsify lipids while the pancreas produces enzymes that catabolize starches, disaccharides, proteins, and fats. These digestive juices break down the food particles in the chyme into glucose, triglycerides, and amino acids. Some chemical digestion of food takes place in the duodenum. Absorption of fatty acids also takes place in the duodenum.

The second part of the small intestine is called the **jejunum**, shown in Figure 5.19. Here, hydrolysis of

nutrients is continued while most of the carbohydrates and amino acids are absorbed through the intestinal lining. The bulk of chemical digestion and nutrient absorption occurs in the jejunum.

The **ileum**, also illustrated in [Figure 5.19](#) is the last part of the small intestine and here the bile salts and vitamins are absorbed into bloodstream. The undigested food is sent to the colon from the ileum via peristaltic movements of the muscle. The ileum ends and the large intestine begins at the ileocecal valve. The vermiform, "worm-like," appendix is located at the ileocecal valve. The appendix of humans secretes no enzymes and has an insignificant role in immunity.

Large intestine

The **large intestine**, illustrated in Figure 5.21, reabsorbs the water from the undigested food material and processes the waste material. The human large intestine is much smaller in length compared to the small intestine but larger in diameter. It has three parts: the cecum, the colon, and the rectum. The cecum joins the ileum to the colon and is the receiving pouch for the waste matter. The colon is home to many bacteria or "intestinal flora" that aid in the digestive processes. The colon can be divided into four regions, the ascending colon, the transverse colon, the descending colon and the sigmoid colon. The main functions of the colon are to extract the water and mineral salts from undigested food, and to store waste material. Carnivorous mammals have a shorter large intestine compared to herbivorous mammals due to their diet.
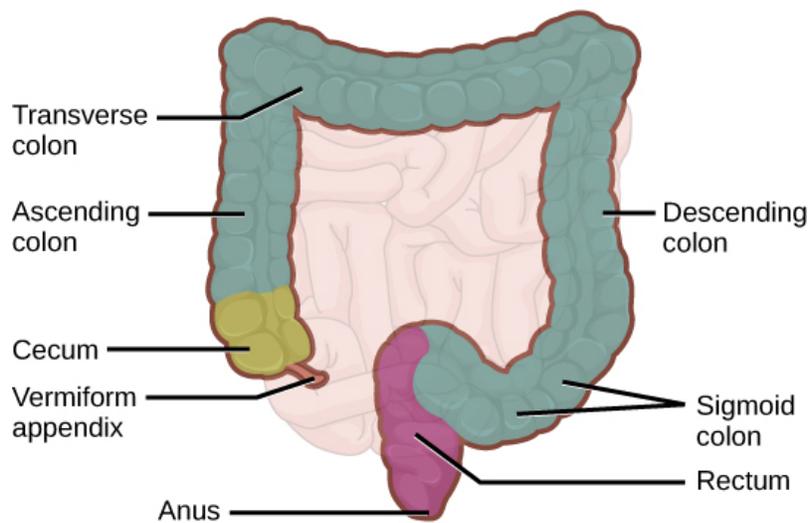

Figure 5.21. The large intestine reabsorbs water from undigested food and stores waste material until it is eliminated.

Rectum and anus

The rectum is the terminal end of the large intestine, as shown in [Figure 5.21](#). The primary role of the rectum is to store the feces until defecation. The feces are propelled using peristaltic movements during elimination. The *anus* is an opening at the far-end of the digestive tract and is the exit point for the waste material. Two sphincters between the rectum and anus control elimination: the inner sphincter is involuntary and the outer sphincter is voluntary.

Accessory organs

The organs discussed above are the organs of the digestive tract through which food passes. Accessory organs are organs that add secretions (enzymes) that catabolize food into nutrients. Accessory organs include salivary glands, the liver, the pancreas, and the gallbladder. The liver, pancreas, and gallbladder are regulated by hormones in response to the food consumed.

The **liver** is the largest internal organ in humans and it plays a very important role in the digestion of fats and detoxifying blood. The liver produces bile, a digestive juice that is required for the breakdown of fatty components of the food in the duodenum. The liver also processes the vitamins and fats and synthesizes many plasma proteins.

The **pancreas** is another important gland that secretes digestive juices. The chyme produced from the stomach is highly acidic in nature; the pancreatic juices contain high levels of bicarbonate, an alkali that neutralizes the acidic chyme. Additionally, the pancreatic juices contain a large variety of enzymes that are required for the digestion of protein and carbohydrates.

The **gallbladder** is a small organ that aids the liver by storing bile and concentrating bile salts. When chyme containing fatty acids enters the duodenum, the bile is secreted from the gallbladder into the duodenum.

Question 5.6

Compare and contrast components of the digestive system among nematodes, rabbit, and chicken?

Question 5.7

If a structure contains thin epithelium, hydrogen pump, and many folds, what could it be good for during digestion? Explain your reasoning.

Question 5.8

How do plants obtain nutrients?

5.3 – Nutrient Transport and Energy Metabolism

5.3. Summarize chemical digestion of carbohydrates, proteins and fats and energy metabolism.

Food energy and ATP

Animals need food to obtain energy and maintain homeostasis. Homeostasis is the ability of a system to maintain a stable internal environment even in the face of external changes to the environment. For example, the normal body temperature of humans is 37°C (98.6°F). Humans maintain this temperature even when the external temperature is hot or cold. It takes energy to maintain this body temperature, and animals obtain this energy from food.

The primary source of energy for animals is carbohydrates, mainly glucose. Glucose is called the body's fuel. The digestible carbohydrates in an animal's diet are converted to glucose molecules through a series of catabolic chemical reactions.

Adenosine triphosphate, or ATP, is the primary energy currency in cells; ATP stores energy in phosphate ester bonds. ATP releases energy when the phosphodiester bonds are broken and ATP is converted to ADP and a phosphate group. ATP is produced by the oxidative reactions in the cytoplasm and mitochondrion of the cell, where carbohydrates, proteins, and fats undergo a series of metabolic reactions collectively called cellular respiration. For example, glycolysis is a series of reactions in which glucose is converted to pyruvic acid and some of its chemical potential energy is transferred to NADH and ATP.

ATP is required for all cellular functions. It is used to build the organic molecules that are required for cells and tissues; it provides energy for muscle contraction and for the transmission of electrical signals in the nervous system. When the amount of ATP is available in excess of the body's requirements, the liver uses the excess ATP and excess glucose to produce molecules called glycogen. Glycogen is a polymeric form of glucose and is stored in the liver and skeletal muscle cells. When blood sugar drops, the liver releases glucose from stores of glycogen. Skeletal muscle converts

glycogen to glucose during intense exercise. The process of converting glucose and excess ATP to glycogen and the storage of excess energy is an evolutionarily important step in helping animals deal with mobility, food shortages, and famine.

Digestive system processes

Obtaining nutrition and energy from food is a multi-step process. For true animals, the first step is ingestion, the act of taking in food. This is followed by digestion, absorption, and elimination. In the following sections, each of these steps will be discussed in detail.

Ingestion

The large molecules found in intact food cannot pass through the cell membranes. Food needs to be broken into smaller particles so that animals can harness the nutrients and organic molecules. The first step in this process is *ingestion.* Ingestion is the process of taking in food through the mouth. In vertebrates, the teeth, saliva, and tongue play important roles in mastication (preparing the food into bolus). While the food is being mechanically broken down, the enzymes in saliva begin to chemically process the food as well. The combined action of these processes modifies the food from large particles to a soft mass that can be swallowed and can travel the length of the esophagus.

Digestion and absorption

Digestion is the mechanical and chemical break down of food into small organic fragments. It is important to break down macromolecules into smaller fragments that are of suitable size for absorption across the digestive epithelium. Large, complex molecules of proteins, polysaccharides, and lipids must be reduced to simpler particles such as simple sugar before they can be absorbed by the digestive epithelial cells. Different organs play specific roles in the digestive process. The animal diet needs carbohydrates, protein, and fat, as well as vitamins and inorganic components for nutritional balance. How each of these components is digested is discussed in the following sections.

Carbohydrates

The digestion of carbohydrates begins in the mouth. The salivary enzyme amylase begins the

breakdown of food starches into maltose, a disaccharide. As the bolus of food travels through the esophagus to the stomach, no significant digestion of carbohydrates takes place. The esophagus produces no digestive enzymes but does produce mucous for lubrication. The acidic environment in the stomach stops the action of the amylase enzyme.

The next step of carbohydrate digestion takes place in the duodenum. Recall that the chyme from the stomach enters the duodenum and mixes with the digestive secretion from the pancreas, liver, and gallbladder. Pancreatic juices also contain amylase, which continues the breakdown of starch and glycogen into maltose, a disaccharide. The disaccharides are broken down into monosaccharides by enzymes called maltases, sucrases, and **lactases**, which are also present in the brush border of the small intestinal wall. Maltase breaks down maltose into glucose. Other disaccharides, such as sucrose and lactose are broken down by sucrase and lactase, respectively. Sucrase breaks down sucrose (or "table sugar") into glucose and fructose, and lactase breaks down lactose (or "milk sugar") into glucose and galactose. The monosaccharides (glucose) thus produced are absorbed and then can be used in metabolic pathways to harness energy. The monosaccharides are transported across the intestinal epithelium into the bloodstream to be transported to the different cells in the body. The steps in carbohydrate digestion are summarized in Figure 5.22 and Table 5.1.
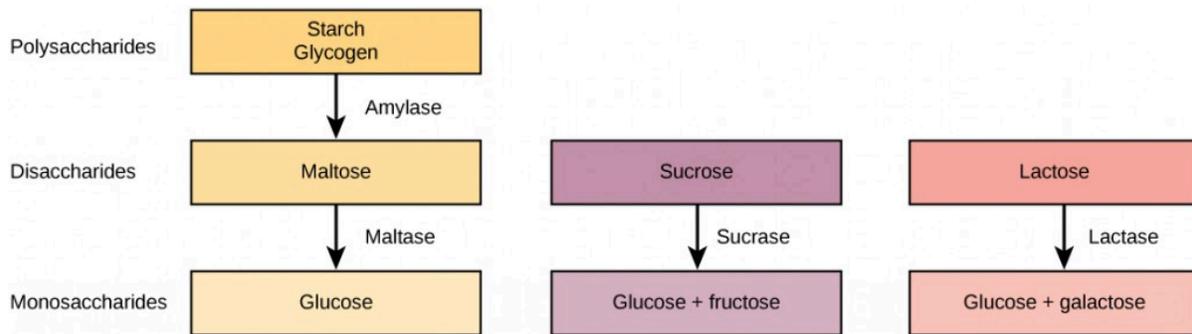
Figure 5.22. Digestion of carbohydrates is performed by several enzymes. Starch and glycogen are broken down into glucose by amylase and maltase. Sucrose (table sugar) and lactose (milk sugar) are broken down by sucrase and lactase, respectively.

Table 5.1. Digestion of carbohydrates

Enzyme	Produced By	Site of Action	Substrate Acting On	End Products
Salivary amylase	Salivary glands	Mouth	Polysaccharides (Starch)	Disaccharides (maltose), oligosaccharides
Pancreatic amylase	Pancreas	Small intestine	Polysaccharides (starch)	Disaccharides (maltose), monosaccharides
Oligosaccharidases	Lining of the intestine; brush border membrane	Small intestine	Disaccharides	Monosaccharides (e.g., glucose, fructose, galactose)

Protein

A large part of protein digestion takes place in the stomach. The enzyme pepsin plays an important role in the digestion of proteins by breaking down the intact protein to peptides, which are short chains of four to nine amino acids. In the duodenum, other enzymes—**trypsin, elastase**, and **chymotrypsin**—act on the peptides reducing them to smaller peptides. Trypsin elastase, carboxypeptidase, and chymotrypsin are produced by the pancreas and released into the duodenum where they act on the chyme. The further breakdown of peptides to single amino acids is aided by enzymes called peptidases (those that break down peptides). Specifically, **carboxypeptidase, dipeptidase**, and **aminopeptidase** play important roles in reducing the peptides to free amino acids. The amino acids are absorbed into the bloodstream through the small intestines. The steps in protein digestion are summarized in Figure 5.23 and Table 5.2.

The liver regulates distribution of amino acids to the rest of the body.

Liver

In the stomach, pepsin breaks down proteins into fragments, called peptides.

Stomach

Pancreas

Amino acids are absorbed from the small intestine into the blood stream.

Protein-digesting enzymes are secreted from the pancreas into the small intestine.

In the small intestine, a variety of enzymes break large peptides into smaller peptides, and then into individual amino acids.

A small amount of dietary protein is lost in the feces.

Figure 5.23. Protein digestion is a multistep process that begins in the stomach and continues through the intestines.

Table 5.2. Digestion of proteins.

Enzyme	Produced By	Site of Action	Substrate Acting On	End Products
Pepsin	Stomach chief cells	Stomach	Proteins	Peptides
• Trypsin • Elastase Chymotrypsin	Pancreas	Small intestine	Proteins	Peptides
Carboxypeptidase	Pancreas	Small intestine	Peptides	Amino acids and pe
• Aminopeptidase • Dipeptidase	Lining of intestine	Small intestine	Peptides	Amino acids

Lipids

Lipid digestion begins in the stomach with the aid of lingual lipase and gastric lipase. However, the bulk of lipid digestion occurs in the small intestine due to pancreatic lipase. When chyme enters the duodenum, the hormonal responses trigger the release of bile, which is produced in the liver and stored in the gallbladder. Bile aids in the digestion of lipids, primarily triglycerides by emulsification. Emulsification is a process in which large lipid globules are broken down into several small lipid globules. These small globules are more widely distributed in the chyme rather than forming large aggregates. Lipids are hydrophobic substances: in the presence of water, they will aggregate to form globules to minimize exposure to water. Bile contains bile salts, which are amphipathic, meaning they contain hydrophobic and hydrophilic parts. Thus, the bile salts hydrophilic side can interface with water on one side and the hydrophobic side interfaces with lipids on the other. By doing so, bile salts emulsify large lipid globules into small lipid globules.

Why is emulsification important for digestion of lipids? Pancreatic juices contain enzymes called lipases (enzymes that break down lipids). If the lipid in the chyme aggregates into large globules, very little surface area of the lipids is available for the lipases to act on, leaving lipid digestion incomplete. By forming an emulsion, bile salts increase the available surface area of the lipids many folds. The pancreatic lipases can then act on the lipids more efficiently and digest them, as detailed in [Figure 5.24](). Lipases break down the lipids into fatty acids and glycerides. These molecules can pass through the plasma membrane of the cell and enter the epithelial cells of the intestinal lining. The bile salts surround long-chain fatty acids and monoglycerides forming tiny spheres called micelles. The micelles move into the brush border of the small intestine absorptive cells where the long-chain fatty acids and monoglycerides diffuse out of the micelles into the absorptive cells leaving the micelles behind in the chyme. The long-chain fatty acids and monoglycerides recombine in the absorptive cells to form triglycerides, which aggregate into globules and become coated with proteins. These large spheres are called **chylomicrons**. Chylomicrons contain triglycerides, cholesterol, and other lipids and have proteins on their surface. The surface is also composed of the hydrophilic phosphate "heads" of phospholipids. Together, they enable the chylomicron to move in an aqueous environment without exposing the lipids to water. Chylomicrons leave the absorptive cells via exocytosis. Chylomicrons enter the lymphatic vessels and then enter the blood in the subclavian vein.
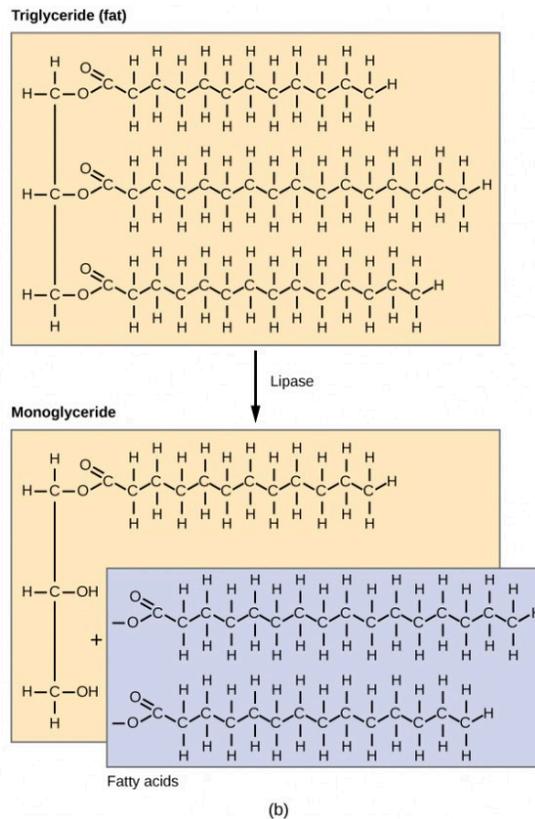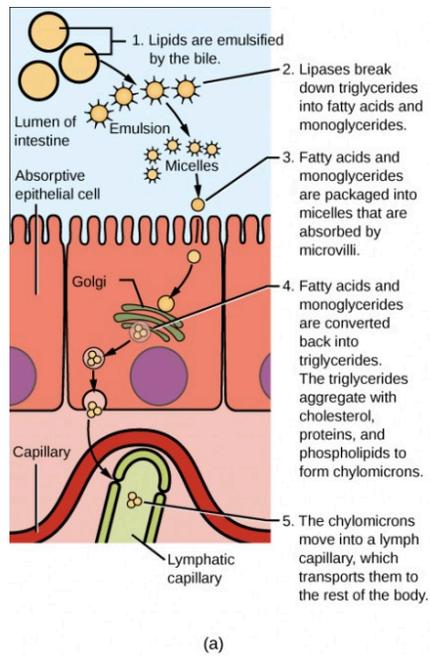
Figure 5.24. Lipids are digested and absorbed in the small intestine.

Vitamins

Vitamins can be either water-soluble or lipid-soluble. Fat-soluble vitamins are absorbed in the same manner as lipids. It is important to consume some amount of dietary lipid to aid the absorption of lipid-soluble vitamins. Water-soluble vitamins can be directly absorbed into the bloodstream from the intestine.

Review Figure 5.25 on your own. You can also use this website, which is an overview of the digestion of protein, fat, and carbohydrates.

• Mechanical digestion (chewing and swallowing)
• Chemical digestion of carbohydrates

• Mechanical digestion (peristaltic mixing and propulsion)
• Chemical digestion of proteins
• Absorption of lipid-soluble substances, such as aspirin

Esophagus

Liver

Gallbladder

Pylorus

Pancreas

• Mechanical digestion (mixing and propulsion, primarily by segmentation)
• Chemical digestion of carbohydrates, lipids, proteins, and nucleic acids
• Absorption of peptides, amino acids, glucose, fructose, lipids, water, minerals, and vitamins

• Mechanical digestion (segmental mixing, mass movement for propulsion)
• No chemical digestion except by bacteria
• Absorption of ions, water, minerals, vitamins, and small organic molecules produced by bacteria

Rectum

Anal sphincters

Figure 5.25. Mechanical and chemical digestion of food takes place in many steps, beginning in the mouth and ending in the rectum.

Question 5.8

Which of the following statements about digestive processes is true?
a. Amylase, maltase, and lactase in the mouth digest carbohydrates.
b. Trypsin and lipase in the stomach digest protein.
c. Bile emulsifies lipids in the small intestine.
d. No food is absorbed until the small intestine.

Elimination

The final step in digestion is the elimination of undigested food content and waste products. The undigested food material enters the colon, where most of the water is reabsorbed. Recall that the colon is also home to the microflora called "intestinal flora" that aid in the digestion process. The semi-solid waste is moved through the colon by peristaltic movements of the muscle and is stored in the rectum. As the rectum expands in response to storage of fecal matter, it triggers the neural signals required to set up the urge to eliminate. The solid waste is eliminated through the anus using peristaltic movements of the rectum.

Diarrhea and constipation are some of the most common health concerns that affect digestion. Constipation is a condition where the feces are hardened because of excess water removal in the colon. In contrast, if enough water is not removed from the feces, it results in diarrhea. Many bacteria, including the ones that cause cholera, affect the proteins involved in water reabsorption in the colon and result in excessive diarrhea.

Emesis, or vomiting, is the elimination of food by forceful expulsion through the mouth. It is often in response to an irritant that affects the digestive tract, including but not limited to viruses, bacteria, emotions, sights, and food poisoning. This forceful expulsion of the food is due to the strong contractions produced by the stomach muscles. The process of emesis is regulated by the medulla.

Question 5.9

Think about the digestion lab you had earlier this week. Summarize digestion of lipids, carbohydrates, and proteins in the digestive system based on your lab observations and results with your lab partners.

Not all animals process different nutrients equally. Here is is a video about hummingbird digestion with a focus on work by Dr. Welch at the University of Toronto Scarborough.

Question 5.10

Why and how are hummingbirds and humans different in carbohydrate digestion?

CHAPTER 6 – LOCOMOTION

6.1. Compare carbohydrate absorption in hummingbirds to absorption of this and other nutrients;

6.2. List different types of muscles and briefly explain how they generate locomotion;

6.3. Explain different types of locomotion with reference to environmental limitations and/or hard structure support for muscle function; and,

6.4. Describe steps involved in muscle contraction.

The content of this chapter was adapted from the *Concepts of Biology–1st Canadian Edition* open textbook by Charles Molnar and Jane Gair and *Anatomy and Physiology* open textbook.

6.1 – Fuels for Locomotion

6.1. Compare carbohydrate absorption in hummingbirds to absorption of this and other nutrients.

Nutrition and energy production

Given the diversity of animal life on our planet, it is not surprising that the animal diet would also vary substantially. The animal diet is the source of materials needed for building DNA and other complex molecules needed for growth, maintenance, reproduction; collectively these processes are called biosynthesis. The diet is also the source of materials for ATP production in the cells, which fuels animal locomotion among many other processes. The diet must be balanced to provide the minerals and vitamins that are required for cellular function.

Food requirements

What are the fundamental requirements of the animal diet? The animal diet should be well balanced and provide nutrients required for bodily function and the minerals and vitamins required for maintaining structure and regulation necessary for good health and reproductive capability. These requirements for a human are illustrated graphically in <u>Figure 6.1</u>.
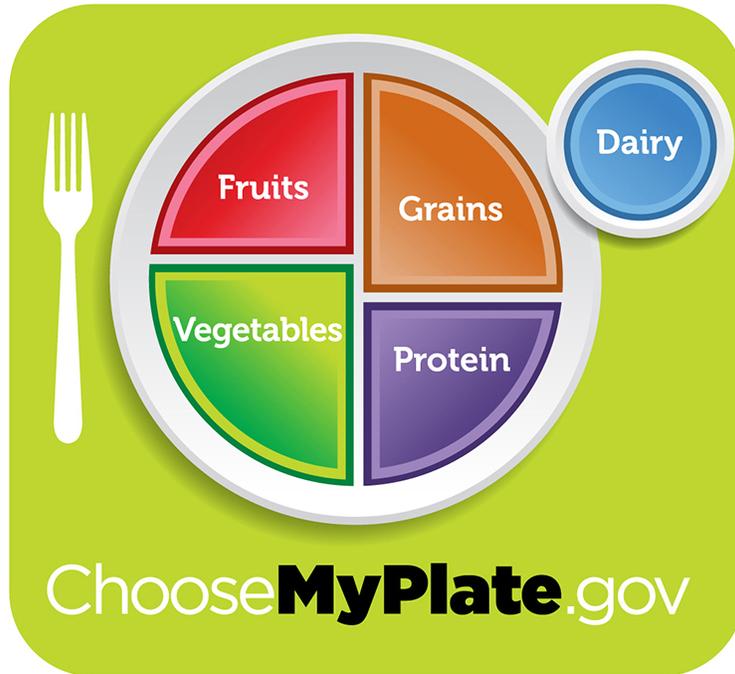
Figure 6.1. For humans, a balanced diet includes fruits, vegetables, grains, and protein. (credit: USDA)

The first step in ensuring that you are meeting the food requirements of your body is an awareness of the food groups and the nutrients they provide. To learn more about each food group and the recommended daily amounts, explore this [interactive site](#) by the United States Department of Agriculture.

Organic precursors

The organic molecules required for building cellular material and tissues must come from food. Carbohydrates or sugars are the primary sources of organic carbons in the animal body. During digestion, digestible carbohydrates are ultimately broken down into glucose and used to provide

energy through metabolic pathways. Complex carbohydrates, including polysaccharides, can be broken down into glucose through biochemical modification; however, humans do not produce the enzyme cellulase and lack the ability to derive glucose from the polysaccharide cellulose. In humans, these molecules provide the fiber required for moving waste through the large intestine and a healthy colon. The intestinal flora in the human gut are able to extract some nutrition from these plant fibers. The excess sugars in the body are converted into glycogen and stored in the liver and muscles for later use. Glycogen stores are used to fuel prolonged exertions, such as long-distance running, and to provide energy during a food shortage. Excess glycogen can be converted to fats, which are stored in the lower layer of the skin of mammals for insulation and energy storage. Excess digestible carbohydrates are stored by mammals in order to survive famine and aid in mobility.

Another important requirement is that of nitrogen. Protein catabolism provides a source of organic nitrogen. Amino acids are the building blocks of proteins and protein breakdown provides amino acids that are used for cellular function. The carbon and nitrogen derived from these become the building block for nucleotides, nucleic acids, proteins, cells, and tissues. Excess nitrogen must be excreted as it is toxic. Fats add flavor to food and promote a sense of satiety or fullness. Fatty foods are also significant sources of energy because one gram of fat contains nine calories. Fats are required in the diet to aid the absorption of fat-soluble vitamins and the production of fat-soluble hormones.

Essential nutrients

While the animal body can synthesize many of the molecules required for the function from the organic precursors, there are some nutrients that need to be consumed from food. These nutrients are termed **essential nutrients**, meaning they must be eaten, and the body cannot produce them.

The omega-3 alpha-linolenic acid and the omega-6 linoleic acid are essential fatty acids needed to make some membrane phospholipids. **Vitamins** are another class of essential organic molecules that are required in small quantities for many enzymes to function and, for this reason, are considered to be co-enzymes. Vitamins can be either water-soluble or lipid-soluble. Fat-soluble vitamins are absorbed in the same manner as lipids. It is important to consume some amount of dietary lipid to aid the absorption of lipid-soluble vitamins. Water-soluble vitamins can be directly absorbed into the bloodstream from the intestine. Absence or low levels of vitamins can have a dramatic effect on health. Both fat-soluble and water-soluble vitamins must be obtained from food. **Minerals** are inorganic essential nutrients that must be obtained from food. Among their many functions, minerals help in structure and regulation and are considered co-factors. Certain amino acids also must be procured from food and cannot be synthesized by the body. These amino acids are the "essential" amino acids. The human body can synthesize only 11 of the 20 required amino acids; the rest must be obtained from food. The essential amino acids are listed in Table 6.1.

You should watch these videos to learn more about water-soluble vitamins, fat-soluble vitamins, and minerals.

Table 6.1. Essential amino acids

Amino acids that must be consumed	Amino acids anabolized by the body
isoleucine	alanine
leucine	selenocysteine
lysine	aspartate
methionine	cysteine
phenylalanine	glutamate
tryptophan	glycine
valine	proline
histidine*	serine
threonine	tyrosine
arginine*	asparagine

*The human body can synthesize histidine and arginine, but not in the quantities required, especially for growing children.

Question 6.1

Summarize roles different vitamins, minerals and essential amino acids play in animal nutrition and energy production.

6.2 – Types of Muscles

![icon]

6.2. List different types of muscles and briefly explain how they generate locomotion.

Types of muscle tissue

Nutrients that animals obtained are used for ATP production, which are fuels for the production of work by muscles, among other physiological processes in the animal bodies. Depending on what type of muscle tissues animals are using, they will use ATP differently to generate work. Overall, muscle cells are specialized for contraction. Muscles allow for motions such as walking, and they also facilitate bodily processes such as respiration and digestion. The body contains three types of muscle tissue: skeletal muscle, cardiac muscle, and smooth muscle (Figure 6.2).

Skeletal muscle   Smooth muscle   Cardiac muscle
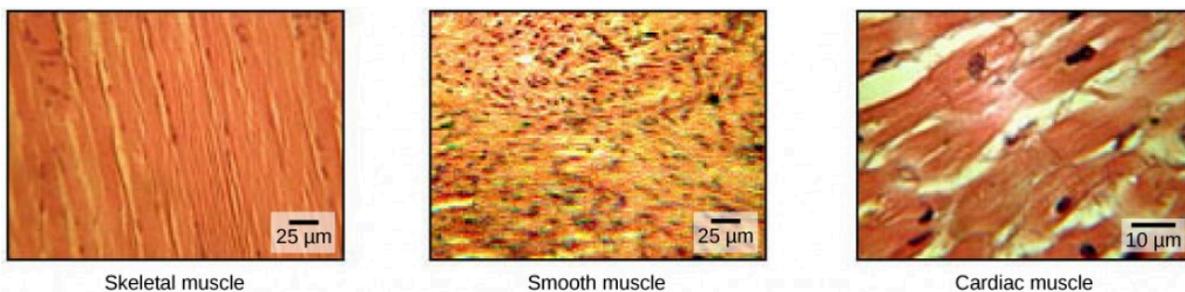
Figure 6.2. The body contains three types of muscle tissue: skeletal muscle, smooth muscle, and cardiac muscle, visualized here using light microscopy. Smooth muscle cells are short, tapered at each end, and have only one plump nucleus in each. Cardiac muscle cells are branched and striated, but short. The cytoplasm may branch, and they have one nucleus in the center of the cell. (credit: modification of work by NCI, NIH; scale-bar data from Matt Russell)

Skeletal muscle tissue forms skeletal muscles, which attach to bones or skin and control locomotion and any movement that can be consciously controlled. Because it can be controlled by thought, skeletal muscle is also called voluntary muscle. Skeletal muscles are long and cylindrical in appearance; when viewed under a microscope, skeletal muscle tissue has a striped or striated appearance. The striations are caused by the regular arrangement of contractile proteins (actin and

myosin). **Actin** is a globular contractile protein that interacts with **myosin** for muscle contraction. Skeletal muscle also has multiple nuclei present in a single cell.

Smooth muscle tissue occurs in the walls of hollow organs such as the intestines, stomach, and urinary bladder, and around passages such as the respiratory tract and blood vessels. Smooth muscle has no striations, is not under voluntary control, has only one nucleus per cell, is tapered at both ends, and is called involuntary muscle.

Cardiac muscle tissue is only found in the heart, and cardiac contractions pump blood throughout the body and maintain blood pressure. Like skeletal muscle, cardiac muscle is striated, but unlike skeletal muscle, cardiac muscle cannot be consciously controlled and is called involuntary muscle. It has one nucleus per cell, is branched, and is distinguished by the presence of intercalated disks.

Skeletal muscle fiber structure

Each skeletal muscle fiber is a skeletal muscle cell. These cells are incredibly large, with diameters of up to 100 μm and lengths of up to 30 cm. The plasma membrane of a skeletal muscle fiber is called the **sarcolemma**. The sarcolemma is the site of action potential conduction, which triggers muscle contraction. Within each muscle fiber are **myofibrils**—long cylindrical structures that lie parallel to the muscle fiber. Myofibrils run the entire length of the muscle fiber, and because they are only approximately 1.2 μm in diameter, hundreds to thousands can be found inside one muscle fiber. They attach to the sarcolemma at their ends, so that as myofibrils shorten, the entire muscle cell contracts ([Figure 6.3](#)).
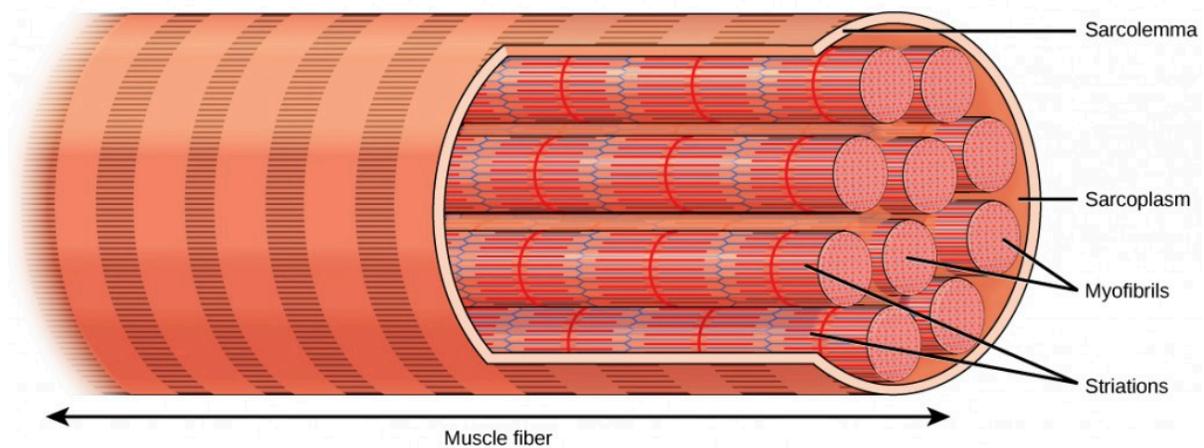
Figure 6.3. A skeletal muscle cell is surrounded by a plasma membrane called the sarcolemma with a cytoplasm called the sarcoplasm. A muscle fiber is composed of many fibrils, packaged into orderly units.

The striated appearance of skeletal muscle tissue is a result of repeating bands of the proteins actin and myosin that are present along the length of myofibrils. Dark A bands and light I bands repeat along myofibrils, and the alignment of myofibrils in the cell causes the entire cell to appear striated or banded.

Each I band has a dense line running vertically through the middle called a Z disc or Z line. The Z discs mark the border of units called **sarcomeres**, which are the functional units of skeletal muscle. One sarcomere is the space between two consecutive Z discs and contains one entire A band and two halves of an I band, one on either side of the A band. A myofibril is composed of many sarcomeres running along its length, and as the sarcomeres individually contract, the myofibrils and muscle cells shorten (Figure 6.4).
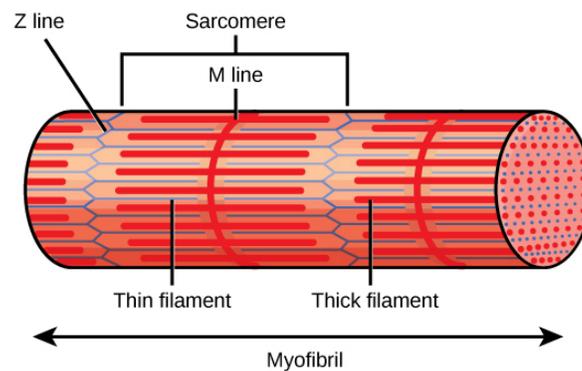
Figure 6.4. A sarcomere is a region from one Z line to the next Z line. Many sarcomeres are present in a myofibril, resulting in the striation pattern characteristic of skeletal muscle.

Myofibrils are composed of smaller structures called **myofilaments**. There are two main types of filaments: thick filaments and thin filaments; each has different compositions and locations. **Thick filaments** occur only in the A band of a myofibril. **Thin filaments** attach to a protein in the Z disc called alpha-actinin and occur across the entire length of the I band and partway into the A band. The region at which thick and thin filaments overlap has a dense appearance, as there is little space between the filaments. Thin filaments do not extend all the way into the A bands, leaving a central region of the A band that only contains thick filaments. This central region of the A band looks slightly lighter than the rest of the A band and is called the H zone. The middle of the H zone has a vertical line called the M line, at which accessory proteins hold together thick filaments. Both the Z disc and the M line hold myofilaments in place to maintain the structural arrangement and layering of the myofibril. Myofibrils are connected to each other by intermediate, or desmin, filaments that attach to the Z disc.

Thick and thin filaments are themselves composed of proteins. Thick filaments are composed of the protein myosin. The tail of a myosin molecule connects with other myosin molecules to form the

central region of a thick filament near the M line, whereas the heads align on either side of the thick filament where the thin filaments overlap. The primary component of thin filaments is the actin protein. Two other components of the thin filament are tropomyosin and troponin. Actin has binding sites for myosin attachment. Strands of tropomyosin block the binding sites and prevent actin-myosin interactions when the muscles are at rest. Troponin consists of three globular subunits. One subunit binds to tropomyosin, one subunit binds to actin, and one subunit binds Ca2+ ions.

View this animation showing the organization of muscle fibers.

In addition to three different types of muscles discussed above, there are also differences in skeletal muscle function. Different skeletal muscles in animals are described as white and red muscle. These different types of skeletal muscles are recruited depending on whether a fast and short versus steady and prolonged locomotion is needed by the animal.

You should read this article about red and white muscles in fish in order to be able to answer the questions at the end of the chapter.

Based on the article about swimming in fish you read (linked above), answer the following questions.

Question 6.2

Describe the differences between white and red muscle.

Based on the article about swimming in fish you read (linked above), answer the following questions.

Question 6.3

Where is the most of the red muscle found in the fish studied in the above article and what is the functional significance of this muscle location? Hint: review Figures 5 and 6 of the article.

Question 6.4.

Which is more efficient – red or white muscle? Why?

6.3 – Types of Locomotion

6.3. Explain different types of locomotion with reference to environmental limitations and/or hard structure support for muscle function.

The muscular and skeletal systems provide support to the body and allow for movement. The bones of the skeleton protect the body's internal organs and support the weight of the body. The muscles of the muscular system contract and pull on the bones, allowing for movements as diverse as standing, walking, running, and grasping items. We will focus on the connection between human skeletal and muscular system to illustrate how these two organ systems coordinate to generate locomotion.

Human skeletal system

The human skeleton is an endoskeleton that consists of 206 bones in the adult. An endoskeleton develops within the body rather than outside like the exoskeleton of insects. The skeleton has five main functions: providing support to the body, storing minerals and lipids, producing blood cells, protecting internal organs, and allowing for movement. The skeletal system in vertebrates is divided into the axial skeleton (which consists of the skull, vertebral column, and rib cage), and the appendicular skeleton (which consists of limb bones, the pectoral or shoulder girdle, and the pelvic girdle).

Explore the human skeleton by viewing the following video with digital 3D sculpturing.

The axial skeleton forms the central axis of the body and includes the bones of the skull, ossicles of the middle ear, hyoid bone of the throat, vertebral column, and the thoracic cage (rib cage) (Figure 6.5).
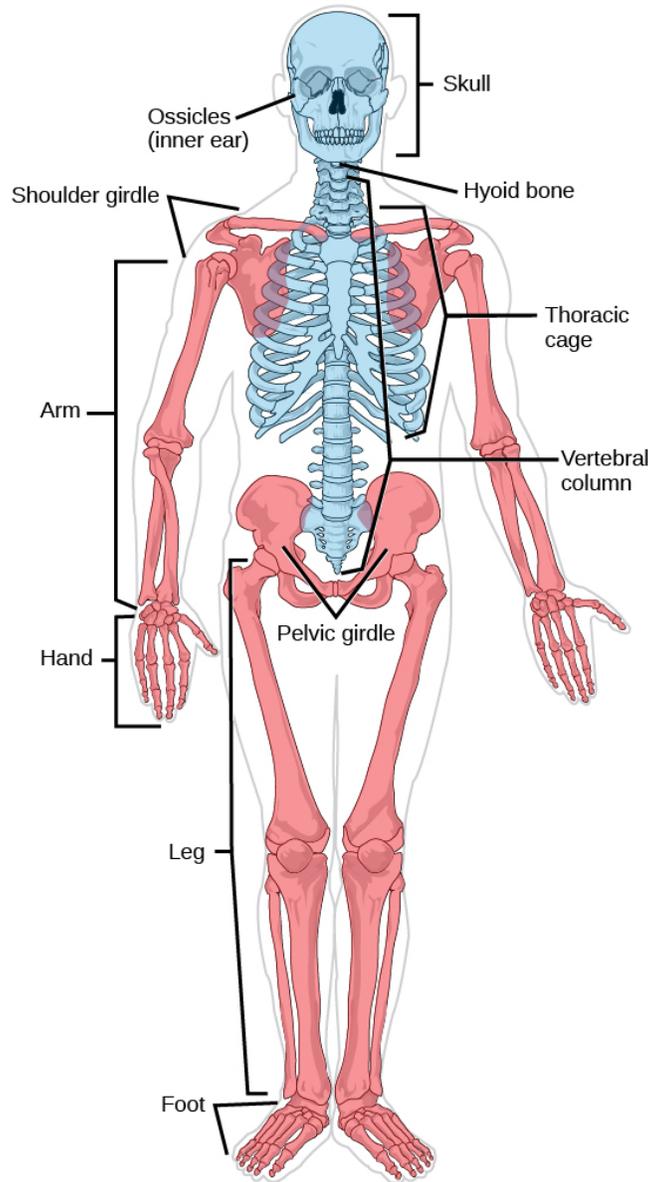
Figure 6.5. The axial skeleton, shown in blue, consists of the bones of the skull, ossicles of the middle ear, hyoid bone, vertebral column, and thoracic cage. The appendicular skeleton, shown in red, consists of the bones of the pectoral limbs, pectoral girdle, pelvic limb, and pelvic girdle. (credit: modification of work by Mariana Ruiz Villareal)

The bones of the skull support the structures of the face and protect the brain. The skull consists of cranial bones and facial bones. The cranial bones form the cranial cavity, which encloses the brain and serves as an attachment site for muscles of the head and neck. In the adult, they are tightly jointed with connective tissue and adjoining bones do not move.

The auditory ossicles of the middle ear transmit sounds from the air as vibrations to the fluid-filled cochlea. The auditory ossicles consist of two malleus (hammer) bones, two incus (anvil) bones, and two stapes (stirrups), one on each side. Facial bones provide cavities for the sense organs (eyes, mouth, and nose), and serve as attachment points for facial muscles.

The hyoid bone lies below the mandible in the front of the neck. It acts as a movable base for the tongue and is connected to muscles of the jaw, larynx, and tongue. The mandible forms a joint with the base of the skull. The mandible controls the opening to the mouth and hence, the airway and gut.

The vertebral column, or spinal column, surrounds and protects the spinal cord, supports the head, and acts as an attachment point for ribs and muscles of the back and neck. It consists of 26 bones: the 24 vertebrae, the sacrum, and the coccyx. Each vertebral body has a large hole in the center through which the spinal cord passes down to the level of the first lumbar vertebra. Below this level, the hole contains spinal nerves which exit between the vertebrae. There is a notch on each side of the hole through which the spinal nerves, can exit from the spinal cord to serve different regions of the body. The vertebral column is approximately 70 cm (28 in) in adults and is curved, which can be seen from a side view.

Intervertebral discs composed of fibrous cartilage lie between adjacent vertebrae from the second cervical vertebra to the sacrum. Each disc helps form a slightly moveable joint and acts as a cushion to absorb shocks from movements such as walking and running.

The thoracic cage, also known as the rib cage consists of the ribs, sternum, thoracic vertebrae, and costal cartilages. The thoracic cage encloses and protects the organs of the thoracic cavity including the heart and lungs. It also provides support for the shoulder girdles and upper limbs and serves as the attachment point for the diaphragm, muscles of the back, chest, neck, and shoulders. Changes in the volume of the thorax enable breathing. The sternum, or breastbone, is a long flat bone located at the anterior of the chest. Like the skull, it is formed from many bones in the embryo, which fuses in the adult. The ribs are 12 pairs of long curved bones that attach to the thoracic vertebrae and curve toward the front of the body, forming the ribcage. Costal cartilages connect the anterior ends of most ribs to the sternum.

The appendicular skeleton is composed of the bones of the upper and lower limbs. It also includes the pectoral, or shoulder girdle, which attaches the upper limbs to the body, and the pelvic girdle, which attaches the lower limbs to the body (Figure 6.5).

The pectoral girdle bones transfer force generated by muscles acting on the upper limb to the thorax. It consists of the clavicles (or collarbones) in the anterior, and the scapulae (or shoulder blades) in the posterior.

The upper limb contains bones of the arm (shoulder to elbow), the forearm, and the hand. The humerus is the largest and longest bone of the upper limb. It forms a joint with the shoulder and with the forearm at the elbow. The forearm extends from the elbow to the wrist and consists of two bones. The hand includes the bones of the wrist, the palm, and the bones of the fingers.

The pelvic girdle attaches to the lower limbs of the axial skeleton. Since it is responsible for bearing the weight of the body and for locomotion, the pelvic girdle is securely attached to the axial skeleton by strong ligaments. It also has deep sockets with robust ligaments that securely attach to the femur. The pelvic girdle is mainly composed of two large hip bones. The hip bones join together in the anterior of the body at a joint called the pubic symphysis and with the bones of the sacrum at the posterior of the body.

The lower limb consists of the thigh, the leg, and the foot. The bones of the lower limbs are thicker and stronger than the bones of the upper limbs to support the entire weight of the body and the forces from locomotion. The femur, or thighbone, is the longest, heaviest, and strongest bone in the body. The femur and pelvis form the hip joint. At its other end, the femur, along with the shinbone and kneecap, form the knee joint.

Joints and skeletal movement

The point at which two or more bones meet is called a joint, or articulation. Joints are responsible for movement, such as the movement of limbs, and stability, such as the stability found in the bones of the skull.

There are two ways to classify joints: based on their structure or based on their function. The structural classification divides joints into fibrous, cartilaginous, and synovial joints depending on the material composing the joint and the presence or absence of a cavity in the joint. The bones of fibrous joints are held together by fibrous connective tissue. There is no cavity, or space, present between the bones, so most fibrous joints do not move at all, or are only capable of minor movements. The joints between the bones in the skull and between the teeth and the bone of their sockets are examples of fibrous joints ([Figure 6.6a](#)).

Cartilaginous joints are joints in which the bones are connected by cartilage ([Figure 6.6b](#)). An example is found at the joints between vertebrae, the so-called "disks" of the backbone. Cartilaginous joints allow for very little movement.

Synovial joints are the only joints that have a space between the adjoining bones ([Figure 6.6c](#)). This space is referred to as the joint cavity and is filled with fluid. The fluid lubricates the joint, reducing friction between the bones and allowing for greater movement. The ends of the bones are covered with cartilage and the entire joint is surrounded by a capsule. Synovial joints are capable of the greatest movement of the joint types. Knees, elbows, and shoulders are examples of synovial joints.
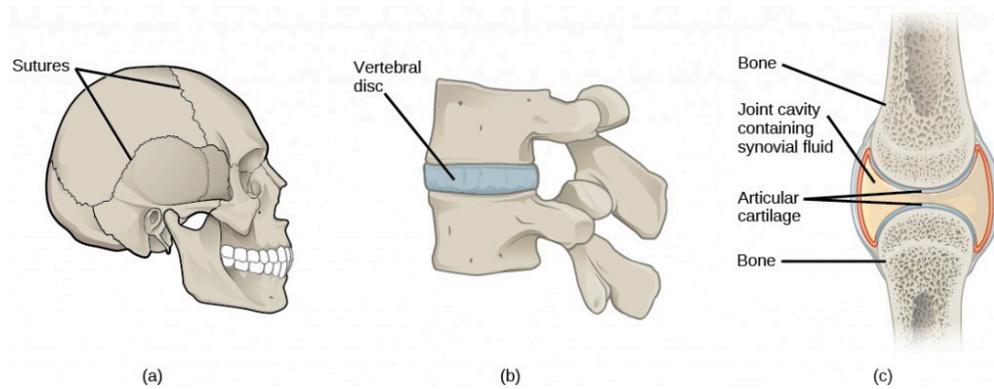
Figure 6.6 (a) Sutures are fibrous joints found only in the skull. (b) Cartilaginous joints are bones connected by cartilage, such as between vertebrae. (c) Synovial joints are the only joints that have a space or "synovial cavity" in the joint.

The wide range of movement allowed by synovial joints produces different types of movements. Angular movements are produced when the angle between the bones of a joint changes. Flexion, or bending, occurs when the angle between the bones decreases. Moving the forearm upward at the elbow is an example of flexion. Extension is the opposite of flexion in that the angle between the bones of a joint increases. Rotational movement is the movement of a bone as it rotates around its own longitudinal axis. Movement of the head as in saying "no" is an example of rotation.

Animal locomotion

Humans are limited to terrestrial locomotion on two limbs or swimming with all four limbs. Other animals explore both the aquatic and aerial realm more extensively. For example, birds are excellent swimmers and fliers. However, some birds are not good at moving in both air and water. One example of birds with limited abilities in flight are penguins and another are members of an extinct group Hesperirnithformes, who, just like penguins, were not able to swim in water and only had limited locomotion on land.

Watch the video about swimming locomotion in extinct Hesperornithiformes to help you understand how skeleton and muscles coordinate to generate locomotion and why some animals may be more limited in locomotion in one type of environment.

Question 6.5

Which mode of locomotion do you think is least efficient in terms of total energy expenditure if you are considering animals of different size?

6.4 – Muscle Contraction

6.4. Describe steps involved in muscle contraction.

Sliding filament model of contraction

For a muscle cell to contract, the sarcomere must shorten. However, thick and thin filaments—the components of sarcomeres—do not shorten. Instead, they slide by one another, causing the sarcomere to shorten while the filaments remain the same length. The sliding filament theory of muscle contraction was developed to fit the differences observed in the named bands on the sarcomere at different degrees of muscle contraction and relaxation. The mechanism of contraction is the binding of myosin to actin, forming cross-bridges that generate filament movement (Figure 6.7).
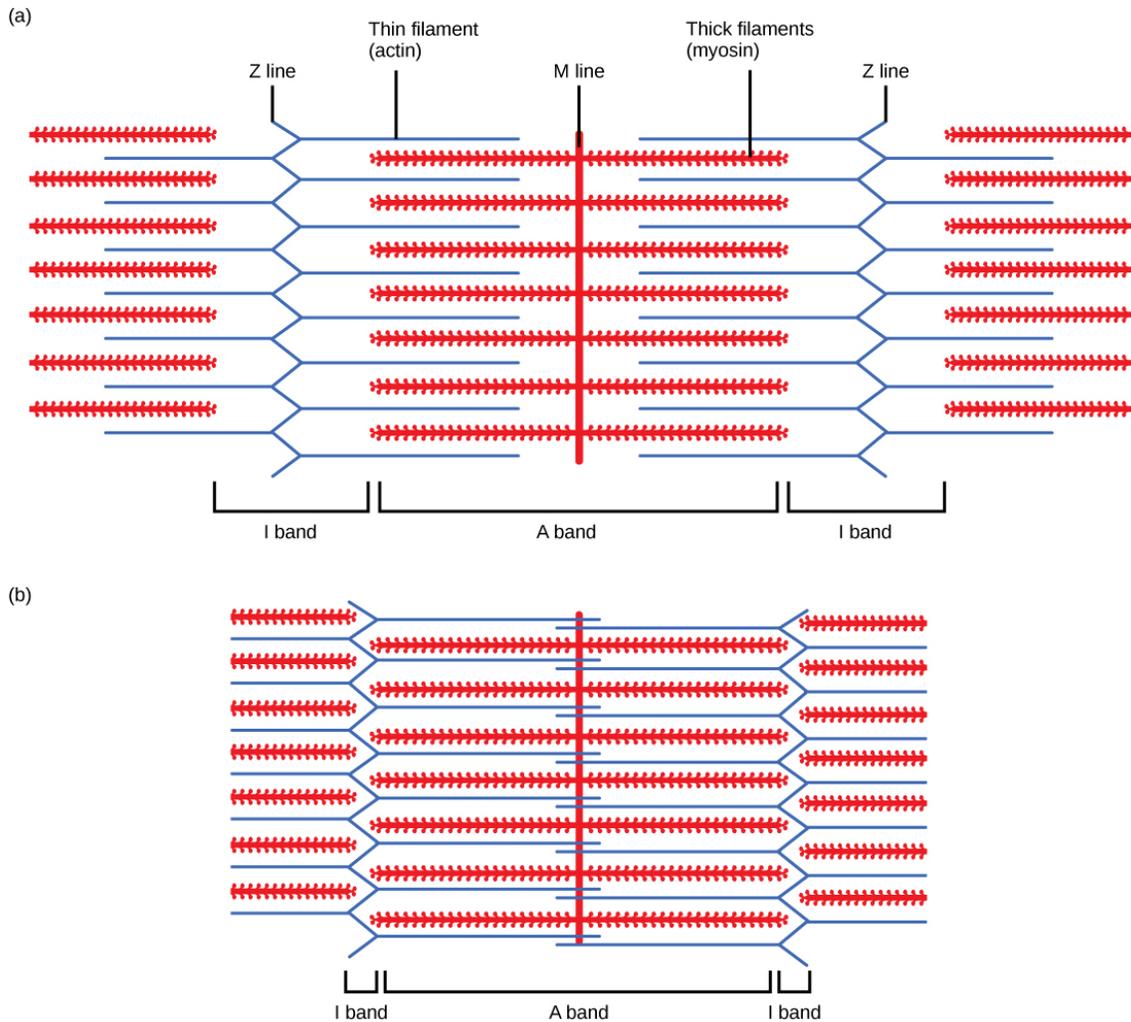
Figure 6.7. When (a) a sarcomere (b) contracts, the Z lines move closer together and the I band gets smaller. The A band stays the same width and, at full contraction, the thin filaments overlap.

When a sarcomere shortens, some regions shorten whereas others stay the same length. A sarcomere is defined as the distance between two consecutive Z discs or Z lines; when a muscle contracts, the distance between the Z discs is reduced. The H zone–the central region of the A zone–contains only thick filaments and is shortened during contraction. The I band contains only thin filaments and also shortens. The A band does not shorten–it remains the same length–but A bands of different sarcomeres move closer together during contraction, eventually disappearing. Thin filaments are pulled by the thick filaments toward the center of the sarcomere until the Z discs approach the thick filaments. The zone of overlap, in which thin filaments and thick filaments occupy the same area, increases as the thin filaments move inward.

ATP and muscle contraction

The motion of muscle shortening occurs as myosin heads bind to actin and pull the actin inwards. This action requires energy, which is provided by ATP. Myosin binds to actin at a binding site on the globular actin protein. Myosin has another binding site for ATP at which enzymatic activity hydrolyzes ATP to ADP, releasing an inorganic phosphate molecule and energy.

ATP binding causes myosin to release actin, allowing actin and myosin to detach from each other. After this happens, the newly bound ATP is converted to ADP and inorganic phosphate, Pi. The enzyme at the binding site on myosin is called ATPase. The energy released during ATP hydrolysis changes the angle of the myosin head into a "cocked" position. The myosin head is then in a position for further movement, possessing potential energy, but ADP and Pi are still attached. If actin binding sites are covered and unavailable, the myosin will remain in the high energy configuration with ATP hydrolyzed but still attached.

If the actin binding sites are uncovered, a cross-bridge will form; that is, the myosin head spans the distance between the actin and myosin molecules. Pi is then released, allowing myosin to expend the stored energy as a conformational change. The myosin head moves toward the M line, pulling the actin along with it. As the actin is pulled, the filaments move approximately 10 nm toward the M line. This movement is called the power stroke, as it is the step at which force is produced. As the actin is pulled toward the M line, the sarcomere shortens and the muscle contracts.

When the myosin head is "cocked," it contains energy and is in a high-energy configuration. This energy is expended as the myosin head moves through the power stroke; at the end of the power stroke, the myosin head is in a low-energy position. After the power stroke, ADP is released; however, the cross-bridge formed is still in place, and actin and myosin are bound together. ATP can then attach to myosin, which allows the cross-bridge cycle to start again and further muscle contraction can occur (Figure 6.8).

Watch this video explaining how a muscle contraction is signaled.

1 The active site on actin is exposed as Ca^{2+} binds troponin.

Troponin Ca^{2+} Tropomyosin
Active site
Myosin head ADP Actin
P

2 The myosin head forms a cross-bridge with actin.

ADP
P

3 During the power stroke, the myosin head bends, and ADP and phosphate are released.

ADP
P

4 A new molecule of ATP attaches to the myosin head, causing the cross-bridge to detach.

ATP

5 ATP hydrolyzes to ADP and phosphate, which returns the myosin to the "cocked" position.

ADP
P

Figure 6.8. The cross–bridge muscle contraction cycle, which is triggered by Ca2+ binding to the actin active site, is shown. With each contraction cycle, actin moves relative to myosin.

Question 6.6

Which of the following statements about muscle contraction is true?
a. The power stroke occurs when ATP is hydrolyzed to ADP and phosphate.
b. The power stroke occurs when ADP and phosphate dissociate from the myosin head.
c. The power stroke occurs when ADP and phosphate dissociate from the actin active site.
d. The power stroke occurs when Ca2+ binds the calcium head.

View this video animation of the cross-bridge muscle contraction.

Regulatory proteins

When a muscle is in a resting state, actin and myosin are separated. To keep actin from binding to the active site on myosin, regulatory proteins block the molecular binding sites. Tropomyosin blocks myosin binding sites on actin molecules, preventing cross-bridge formation and preventing contraction in a muscle without nervous input. Troponin binds to tropomyosin and helps to position it on the actin molecule; it also binds calcium ions.

To enable a muscle contraction, tropomyosin must change conformation, uncovering the myosin-binding site on an actin molecule and allowing cross-bridge formation. This can only happen in the presence of calcium, which is kept at extremely low concentrations in the sarcoplasm. If present, calcium ions bind to troponin, causing conformational changes in troponin that allow tropomyosin to move away from the myosin binding sites on actin. Once the tropomyosin is removed, a cross-bridge can form between actin and myosin, triggering contraction. Cross-bridge cycling continues until Ca2+ ions and ATP are no longer available and tropomyosin again covers the binding sites on actin.

Excitation-contraction coupling

Excitation-contraction coupling is the link (transduction) between the action potential generated in the sarcolemma and the start of a muscle contraction. The trigger for calcium release from the sarcoplasmic reticulum into the sarcoplasm is a neural signal. Each skeletal muscle fiber is controlled by a motor neuron, which conducts signals from the brain or spinal cord to the muscle. The area of the sarcolemma on the muscle fiber that interacts with the neuron is called the motor end plate. The end of the neuron's axon is called the synaptic terminal, and it does not actually contact the motor end plate. A small space called the synaptic cleft separates the synaptic terminal from the motor end plate. Electrical signals travel along the neuron's axon, which branches through the muscle and connects to individual muscle fibers at a neuromuscular junction.

The ability of cells to communicate electrically requires that the cells expend energy to create

an electrical gradient across their cell membranes. This charge gradient is carried by ions, which are differentially distributed across the membrane. Each ion exerts an electrical influence and a concentration influence. Just as milk will eventually mix with coffee without the need to stir, ions also distribute themselves evenly, if they are permitted to do so. In this case, they are not permitted to return to an evenly mixed state.

The sodium-potassium ATPase uses cellular energy to move K+ ions inside the cell and Na+ ions outside. This alone accumulates a small electrical charge, but a big concentration gradient. There is lots of K+ in the cell and lots of Na+ outside the cell. Potassium is able to leave the cell through K+ channels that are open 90% of the time, and it does. However, Na+ channels are rarely open, so Na+remains outside the cell. When K+ leaves the cell, obeying its concentration gradient, that effectively leaves a negative charge behind. So at rest, there is a large concentration gradient for Na+ to enter the cell, and there is an accumulation of negative charges left behind in the cell. This is the resting membrane potential. The potential in this context means a separation of electrical charge that is capable of doing work. It is measured in volts, just like a battery. However, the transmembrane potential is considerably smaller (0.07 V); therefore, the small value is expressed as millivolts (mV) or 70 mV. Because the inside of a cell is negative compared with the outside, a minus sign signifies the excess of negative charges inside the cell, –70 mV.

If an event changes the permeability of the membrane to Na+ ions, they will enter the cell. That will change the voltage. This is an electrical event, called an action potential, that can be used as a cellular signal. Communication occurs between nerves and muscles through neurotransmitters. Neuron action potentials cause the release of neurotransmitters from the synaptic terminal into the synaptic cleft, where they can then diffuse across the synaptic cleft and bind to a receptor molecule on the motor end plate. The motor end plate possesses junctional folds—folds in the sarcolemma that create a large surface area for the neurotransmitter to bind to receptors. The receptors are actually sodium channels that open to allow the passage of Na+ into the cell when they receive neurotransmitter signal.

Acetylcholine (ACh) is a neurotransmitter released by motor neurons that binds to receptors in the motor end plate. Neurotransmitter release occurs when an action potential travels down the motor neuron's axon, resulting in altered permeability of the synaptic terminal membrane and an influx of calcium. The Ca2+ ions allow synaptic vesicles to move to and bind with the presynaptic membrane (on the neuron) and release neurotransmitter from the vesicles into the synaptic cleft. Once released by the synaptic terminal, ACh diffuses across the synaptic cleft to the motor end plate, where it binds with ACh receptors. As a neurotransmitter binds, these ion channels open, and Na+ ions cross the membrane into the muscle cell. This reduces the voltage difference between the inside and outside of the cell, which is called depolarization. As ACh binds at the motor end plate, this depolarization is called an end-plate potential. The depolarization then spreads along the sarcolemma, creating an action potential as sodium channels adjacent to the initial depolarization site detect the change in voltage and open. The action potential moves across the entire cell, creating a wave of depolarization.

ACh is broken down by the enzyme acetylcholinesterase (AChE) into acetyl and choline. AChE resides

in the synaptic cleft, breaking down ACh so that it does not remain bound to ACh receptors, which would cause unwanted extended muscle contraction (Figure 6.9).
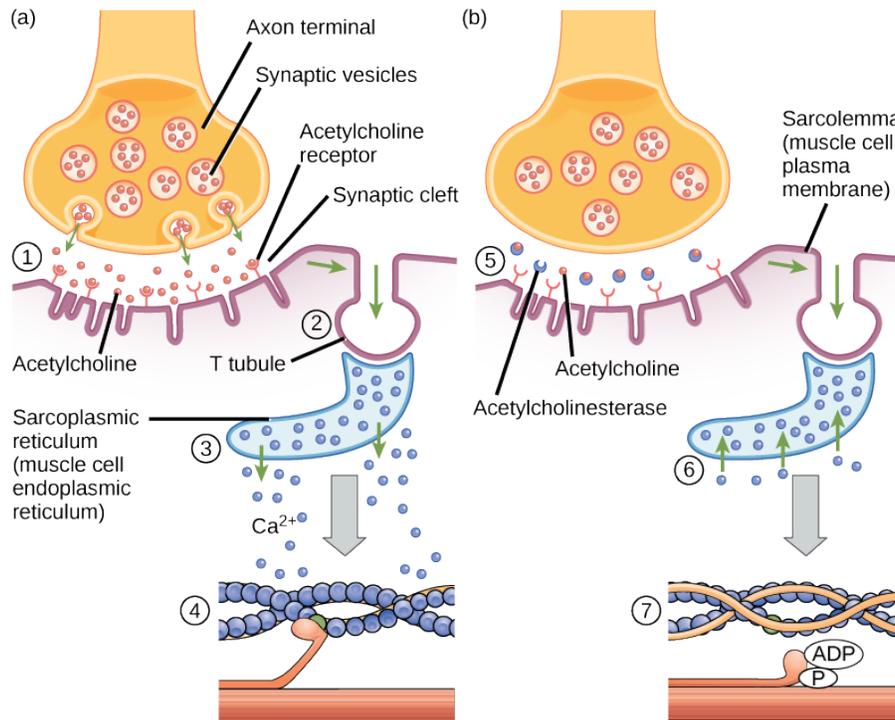
1. Acetylcholine released from the axon terminal binds to receptors on the sarcolemma.
2. An action potential is generated and travels down the T tubule.
3. Ca^{2+} is released from the sarcoplasmic reticulum in response to the change in voltage.
4. Ca^{2+} binds troponin; Cross-bridges form between actin and myosin.

5. Acetylcholinesterase removes acetylcholine from the synaptic cleft.
6. Ca^{2+} is transported back into the sarcoplasmic reticulum.
7. Tropomyosin binds active sites on actin causing the cross-bridge to detach.

Figure 6.9. This diagram shows excitation–contraction coupling in a skeletal muscle contraction. The sarcoplasmic reticulum is a specialized endoplasmic reticulum found in muscle cells.

Question 6.7

The deadly nerve gas Sarin irreversibly inhibits acetylcholinesterase. What effect would Sarin have on muscle contraction?

After depolarization, the membrane returns to its resting state. This is called repolarization, during which voltage-gated sodium channels close. Potassium channels continue at 90% conductance.

Because the plasma membrane sodium–potassium ATPase always transports ions, the resting state (negatively charged inside relative to the outside) is restored. The period immediately following the transmission of an impulse in a nerve or muscle, in which a neuron or muscle cell regains its ability to transmit another impulse, is called the refractory period. During the refractory period, the membrane cannot generate another action potential. The refractory period allows the voltage-sensitive ion channels to return to their resting configurations. The sodium-potassium ATPase continually moves Na+ back out of the cell and K+ back into the cell, and the K+ leaks out leaving negative charge behind. Very quickly, the membrane repolarizes, so that it can again be depolarized.

Control of muscle tension

Neural control initiates the formation of actin-myosin cross-bridges, leading to the sarcomere shortening involved in muscle contraction. These contractions extend from the muscle fiber through connective tissue to pull on bones, causing skeletal movement. The pull exerted by a muscle is called tension, and the amount of force created by this tension can vary. This enables the same muscles to move very light objects and very heavy objects. In individual muscle fibers, the amount of tension produced depends on the cross-sectional area of the muscle fiber and the frequency of neural stimulation.

The number of cross-bridges formed between actin and myosin determines the amount of tension that a muscle fiber can produce. Cross-bridges can only form where thick and thin filaments overlap, allowing myosin to bind to actin. If more cross-bridges are formed, more myosin will pull on actin, and more tension will be produced.

The ideal length of a sarcomere during the production of maximal tension occurs when thick and thin filaments overlap to the greatest degree. If a sarcomere at rest is stretched past an ideal resting length, thick and thin filaments do not overlap to the greatest degree, and fewer cross-bridges can form. This results in fewer myosin heads pulling on actin, and less tension is produced. As a sarcomere is shortened, the zone of overlap is reduced as the thin filaments reach the H zone, which is composed of myosin tails. Because it is myosin heads that form cross-bridges, actin will not bind to myosin in this zone, reducing the tension produced by this myofiber. If the sarcomere is shortened, even more, thin filaments begin to overlap with each other—reducing cross-bridge formation even further and producing even less tension. Conversely, if the sarcomere is stretched to the point at which thick and thin filaments do not overlap at all, no cross-bridges are formed and no tension is produced. This amount of stretching does not usually occur because accessory proteins, internal sensory nerves, and connective tissue oppose extreme stretching.

The primary variable determining force production is the number of myofibers within the muscle that receive an action potential from the neuron that controls that fiber. When using the biceps to pick up a pencil, the motor cortex of the brain only signals a few neurons of the biceps, and only a

few myofibers respond. In vertebrates, each myofiber responds fully if stimulated. When picking up a piano, the motor cortex signals all of the neurons in the biceps and every myofiber participates. This is close to the maximum force the muscle can produce. As mentioned above, increasing the frequency of action potentials (the number of signals per second) can increase the force a bit more, because the tropomyosin is flooded with calcium.

Question 6.8

Teach your peer about the events during muscle contraction, from the arrival of the neural signal to generation of motion powered by the muscle. When you are done, ask your peer what terms or steps you missed or did not explain well. Let your peer fill the gaps. If there were no gaps, your peer can challenge you with some questions about your explanation. Remember that one way that you can test whether you are learning is to be able to transmit your knowledge to another person.

CHAPTER 7 – ELECTRICAL SIGNALS

7.1. Describe membrane potential in terms of variables in Ohm's Law;

7.2. Differentiate between resting, graded and action potential with reference to structures of the neuron as well as ion movements and membrane potentials;

7.3. Explain how sample signals cross the synapse; and,

7.4. Describe the roles of different neuroglia.

The content of this chapter was adapted from the *Concepts of Biology-1st Canadian Edition* open textbook by Charles Molnar and Jane Gair and *Anatomy and Physiology* open textbook.

7.1 – Electrical Potential

7.1. Describe membrane potential in terms of variables in Ohm's Law.

Ohm's Law

What drives current? We can think of various devices–such as batteries, generators, wall outlets, and so on–which are necessary to maintain a current. All such devices create a potential difference and are loosely referred to as voltage sources. When a voltage source is connected to a conductor, it applies a potential difference V that creates an electric field. The electric field in turn exerts force on charges, causing current.

The current that flows through most substances is directly proportional to the voltage V applied to it. The German physicist Georg Simon Ohm (1787–1854) was the first to demonstrate experimentally that the current in a metal wire is *directly proportional to the voltage applied*:
$I \propto V$.

This important relationship is known as Ohm's law. It can be viewed as a cause-and-effect relationship, with voltage the cause and current the effect. This is an empirical law like that for friction–an experimentally observed phenomenon. Such a linear relationship doesn't always occur.

Resistance and Simple Circuits

If voltage drives current, what impedes it? The electric property that impedes current (crudely similar to friction and air resistance) is called resistance R. Collisions of moving charges with atoms and molecules in a substance transfer energy to the substance and limit current. Resistance is defined as inversely proportional to current, or
$I \propto \frac{1}{R}$.

Thus, for example, current is cut in half if resistance doubles. Combining the relationships of current to voltage and current to resistance gives

$$I = \frac{V}{R}.$$

This relationship is also called Ohm's law. Ohm's law in this form really defines resistance for certain materials. Ohm's law (like Hooke's law) is not universally valid. The many substances for which Ohm's law holds are called ohmic. These include good conductors like copper and aluminum, and some poor conductors under certain circumstances. Ohmic materials have a resistance R that is independent of voltage V and current I. An object that has simple resistance is called a *resistor*, even if its resistance is small. The unit for resistance is an ohm and is given the symbol Ω (upper case Greek omega). Rearranging $I = V/R$ gives $R = V/I$, and so the units of resistance are 1 ohm = 1 volt per ampere:

$$1 \ = 1\frac{V}{A}.$$

[Figure 7.1](#) shows the schematic for a simple circuit. A simple circuit has a single voltage source and a single resistor. The wires connecting the voltage source to the resistor can be assumed to have negligible resistance, or their resistance can be included in R.

A simple electric circuit in which a closed path for current to flow is supplied by conductors (usually metal wires) connecting a load to the terminals of a battery, represented by the red parallel lines. The zigzag symbol represents the single resistor and includes any resistance in the connections to the voltage source.
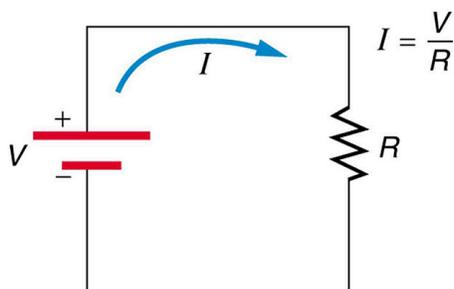
Figure 7.1. Schematic for a simple circuit.

Resistances range over many orders of magnitude. Some ceramic insulators, such as those used to support power lines, have resistances of 10^{12} Ω or more. A dry person may have a hand-to-foot resistance of 10^5 Ω, whereas the resistance of the human heart is about 10^3 Ω. A meter-long piece of large-diameter copper wire may have a resistance of 10^{-5} Ω, and superconductors have no resistance at all (they are non-ohmic). Resistance is related to the shape of an object and the material of which it is composed.

Additional insight is gained by solving $I = V/R$ for V, yielding

$$V = IR.$$

This expression for V can be interpreted as the *voltage drop across a resistor produced by the flow of current I*. The phrase IR *drop* is often used for this voltage. For instance, the headlight in has an IR drop of 12.0 V. If voltage is measured at various points in a circuit, it will be seen to increase at the voltage source and decrease at the resistor. Voltage is similar to fluid pressure. The voltage source is like a pump, creating a pressure difference, causing current–the flow of charge. The resistor is like a pipe that reduces pressure and limits flow because of its resistance. Conservation of energy has important consequences here. The voltage source supplies energy (causing an electric field and a current), and the resistor converts it to another form (such as thermal energy). In a simple circuit (one with a single simple resistor), the voltage supplied by the source equals the voltage drop across the resistor, since $\mathrm{PE} = q\Delta V$, and the same q flows through each. Thus the energy supplied by the voltage source and the energy converted by the resistor are equal ([Figure 7.2](#)).
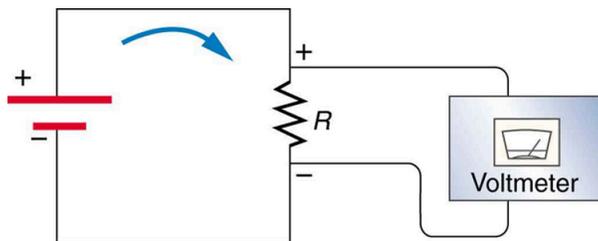
$$V = IR = 18 \text{ V}$$

Figure 7.2. The voltage drop across a resistor in a simple circuit equals the voltage output of the battery.

In a simple electrical circuit, the sole resistor converts energy supplied by the source into another form. Conservation of energy is evidenced here by the fact that all of the energy supplied by the source is converted to another form by the resistor alone. We will find that conservation of energy has other important applications in circuits and is a powerful tool in circuit analysis.

Nerve conduction

Electric currents in the vastly complex system of billions of nerves in our body allow us to sense the world, control parts of our body, and think. These are representative of the three major functions of nerves. First, nerves carry messages from our sensory organs and others to the central nervous system, consisting of the brain and spinal cord. Second, nerves carry messages from the central nervous system to muscles and other organs. Third, nerves transmit and process signals within the central nervous system. The sheer number of nerve cells and the incredibly greater number of connections between them makes this system the subtle wonder that it is. Nerve conduction is a general term for electrical signals carried by nerve cells. It is one aspect of bioelectricity, or electrical effects in and created by biological systems.

Nerve cells, properly called *neurons*, look different from other cells—they have tendrils, some of them many centimeters long, connecting them with other cells (<u>Figure 7.3</u>). Signals arrive at the cell body across *synapses* or through *dendrites*, stimulating the neuron to generate its own signal, sent along its long *axon* to other nerve or muscle cells. Signals may arrive from many other locations and be transmitted to yet others, conditioning the synapses by use, giving the system its complexity and its ability to learn.

A neuron with its dendrites and long axon. Signals in the form of electric currents reach the cell body through dendrites and across synapses, stimulating the neuron to generate its own signal sent down the axon. The number of interconnections can be far greater than shown here.
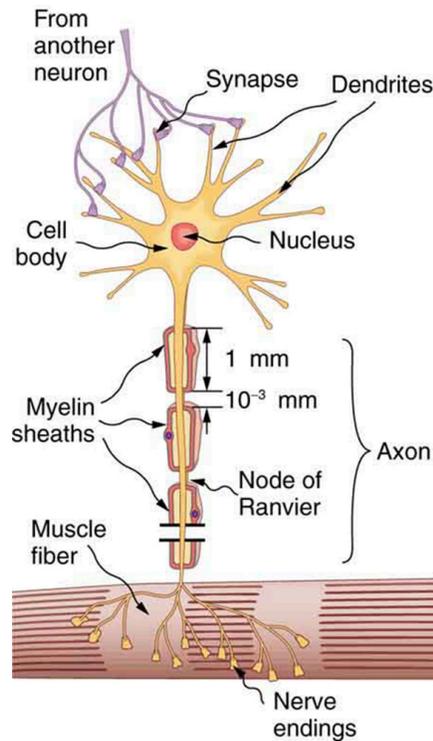
Figure 7.3. The structure of a neuron.

The method by which these electric currents are generated and transmitted is more complex than the simple movement of free charges in a conductor, but it can be understood with principles already discussed in this text. The most important of these are the Coulomb force and diffusion.

<u>Figure 7.4</u> illustrates how a voltage (potential difference) is created across the cell membrane of a neuron in its resting state. This thin membrane separates electrically neutral fluids having differing concentrations of ions, the most important varieties being Na^+, K^+, and Cl^- (these are sodium, potassium, and chlorine ions with single plus or minus charges as indicated). Free ions will diffuse from a region of high concentration to one of low concentration. But the cell membrane is semipermeable, meaning that some ions may cross it while others cannot. In its resting state, the

cell membrane is permeable to K^+ and Cl^-, and impermeable to Na^+. Diffusion of K^+ and Cl^- thus creates the layers of positive and negative charge on the outside and inside of the membrane. The Coulomb force prevents the ions from diffusing across in their entirety. Once the charge layer has built up, the repulsion of like charges prevents more from moving across, and the attraction of unlike charges prevents more from leaving either side. The result is two layers of charge right on the membrane, with diffusion being balanced by the Coulomb force. A tiny fraction of the charges move across and the fluids remain neutral (other ions are present), while a separation of charge and a voltage have been created across the membrane.

The semipermeable membrane of a cell has different concentrations of ions inside and out. Diffusion moves the K^+ and Cl^- ions in the direction shown, until the Coulomb force halts further transfer. This results in a layer of positive charge on the outside, a layer of negative charge on the inside, and thus a voltage across the cell membrane. The membrane is normally impermeable to Na^+.
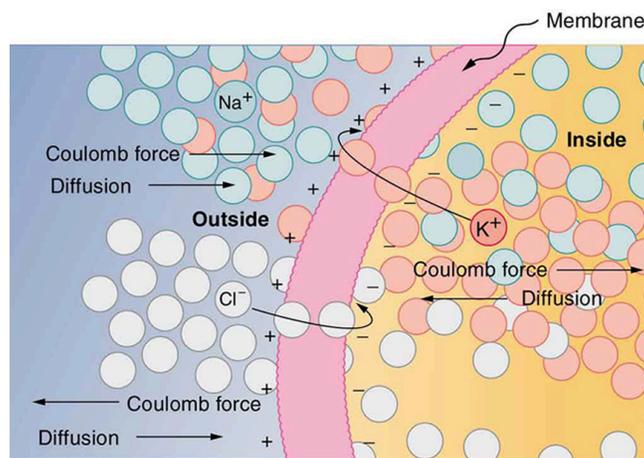
Figure 7.4. The process of how a voltage (potential difference) is created across the cell membrane of a neuron in its resting state.

Question 7.1

Describe electrical signal in a cell using the following terms: resistance, flow, and potential difference.

Question 7.2
Which cells are involved in maintaining electrical potential of an animal cell and how?

7.2 – Resting, Graded and Action Potential

7.2. Differentiate between resting, graded and action potential with reference to structures of the neuron as well as ion movements and membrane potentials.

Neurons

The nervous system of the common laboratory fly, *Drosophila melanogaster*, contains around 100,000 neurons, the same number as a lobster. This number compares to 75 million in the mouse and 300 million in the octopus. A human brain contains around 86 billion neurons. Despite these very different numbers, the nervous systems of these animals control many of the same behaviors—from basic reflexes to more complicated behaviors like finding food and courting mates. The ability of neurons to communicate with each other as well as with other types of cells underlies all of these behaviors.

Most neurons share the same cellular components. But neurons are also highly specialized—different types of neurons have different sizes and shapes that relate to their functional roles.

Parts of a neuron

Like other cells, each neuron has a cell body (or soma) that contains a nucleus, smooth and rough endoplasmic reticulum, Golgi apparatus, mitochondria, and other cellular components. Neurons also contain unique structures, illustrated in [Figure 7.5](#) for receiving and sending the electrical signals that make neuronal communication possible. **Dendrites** are tree-like structures that extend away from the cell body to receive messages from other neurons at specialized junctions called **synapses**. Although some neurons do not have any dendrites, some types of neurons have multiple dendrites. Dendrites can have small protrusions called dendritic spines, which further increase surface area for possible synaptic connections.

Once a signal is received by the dendrite, it then travels passively to the cell body. The cell body contains a specialized structure, the **axon hillock** that integrates signals from multiple synapses and serves as a junction between the cell body and an **axon**. An axon is a tube-like structure that

propagates the integrated signal to specialized endings called **axon terminals**. These terminals in turn synapse on other neurons, muscle, or target organs. Chemicals released at axon terminals allow signals to be communicated to these other cells. Neurons usually have one or two axons, but some neurons, like amacrine cells in the retina, do not contain any axons. Some axons are covered with **myelin**, which acts as an insulator to minimize dissipation of the electrical signal as it travels down the axon, greatly increasing the speed on conduction. This insulation is important as the axon from a human motor neuron can be as long as a meter—from the base of the spine to the toes. The myelin sheath is not actually part of the neuron. Myelin is produced by glial cells. Along the axon there are periodic gaps in the myelin sheath. These gaps are called **nodes of Ranvier** and are sites where the signal is "recharged" as it travels along the axon.

It is important to note that a single neuron does not act alone—neuronal communication depends on the connections that neurons make with one another (as well as with other cells, like muscle cells). Dendrites from a single neuron may receive synaptic contact from many other neurons. For example, dendrites from a Purkinje cell in the cerebellum are thought to receive contact from as many as 200,000 other neurons.
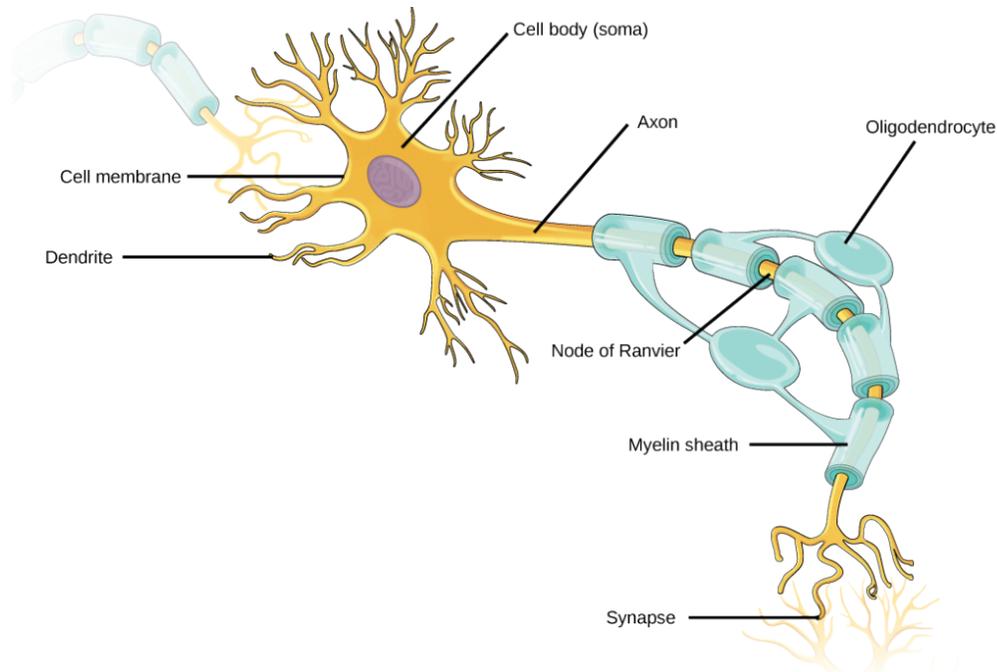
Figure 7.5. Neurons contain organelles common to many other cells, such as a nucleus and mitochondria. They also have more specialized structures, including dendrites and axons.

Question 7.3

Which of the following statements is false?
a. The soma is the cell body of a nerve cell.
b. Myelin sheath provides an insulating layer to the dendrites.
c. Axons carry the signal from the soma to the target.
d. Dendrites carry the signal to the soma.

Types of neurons

There are different types of neurons, and the functional role of a given neuron is intimately dependent on its structure. There is an amazing diversity of neuron shapes and sizes found in different parts of the nervous system (and across species), as illustrated by the neurons shown in Figure 7.6.

(a) Pyramidal cell of the cerebral cortex

(b) Purkinje cell of the cerebellar cortex

(c) Olfactory neurons
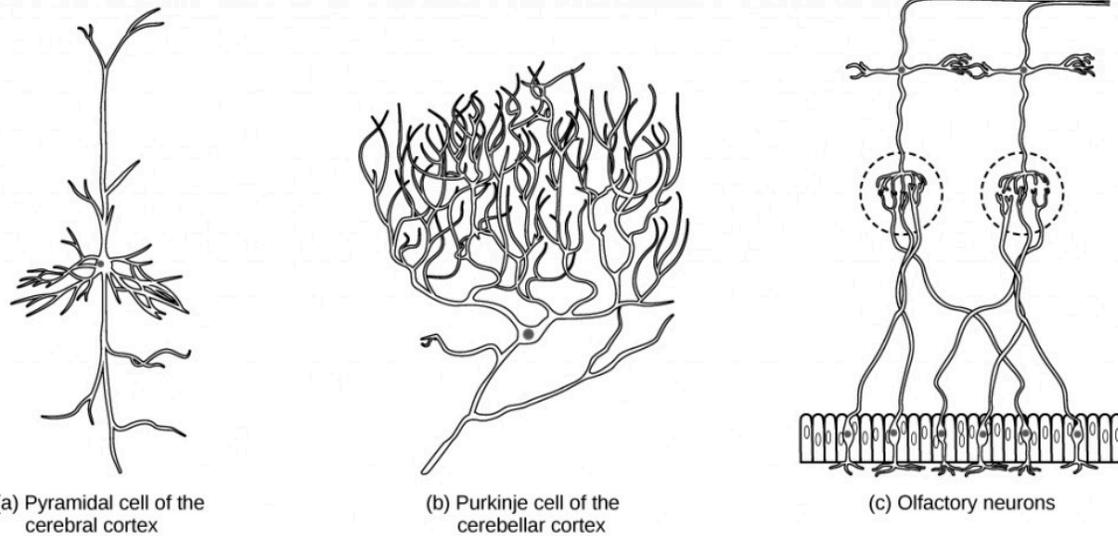
Figure 7.6. There is great diversity in the size and shape of neurons throughout the nervous system. Examples include (a) a pyramidal cell from the cerebral cortex, (b) a Purkinje cell from the cerebellar cortex, and (c) olfactory cells from the olfactory epithelium and olfactory bulb.

While there are many defined neuron cell subtypes, neurons are broadly divided into four basic types: unipolar, bipolar, multipolar, and pseudounipolar. Figure 7.7 illustrates these four basic neuron types. Unipolar neurons have only one structure that extends away from the soma. These neurons are not found in vertebrates but are found in insects where they stimulate muscles or glands. A bipolar neuron has one axon and one dendrite extending from the soma. An example of a bipolar neuron is a retinal bipolar cell, which receives signals from photoreceptor cells that are sensitive to light and transmits these signals to ganglion cells that carry the signal to the brain. Multipolar neurons are the most common type of neuron. Each multipolar neuron contains one axon and multiple dendrites. Multipolar neurons can be found in the central nervous system (brain and spinal cord). An example of a multipolar neuron is a Purkinje cell in the cerebellum, which has many branching dendrites but only one axon. Pseudounipolar cells share characteristics with both unipolar and bipolar cells. A pseudounipolar cell has a single process that extends from the soma, like a unipolar cell, but this process later branches into two distinct structures, like a bipolar cell. Most sensory neurons are pseudounipolar and have an axon that branches into two extensions: one connected to dendrites that receive sensory information and another that transmits this information to the spinal cord.
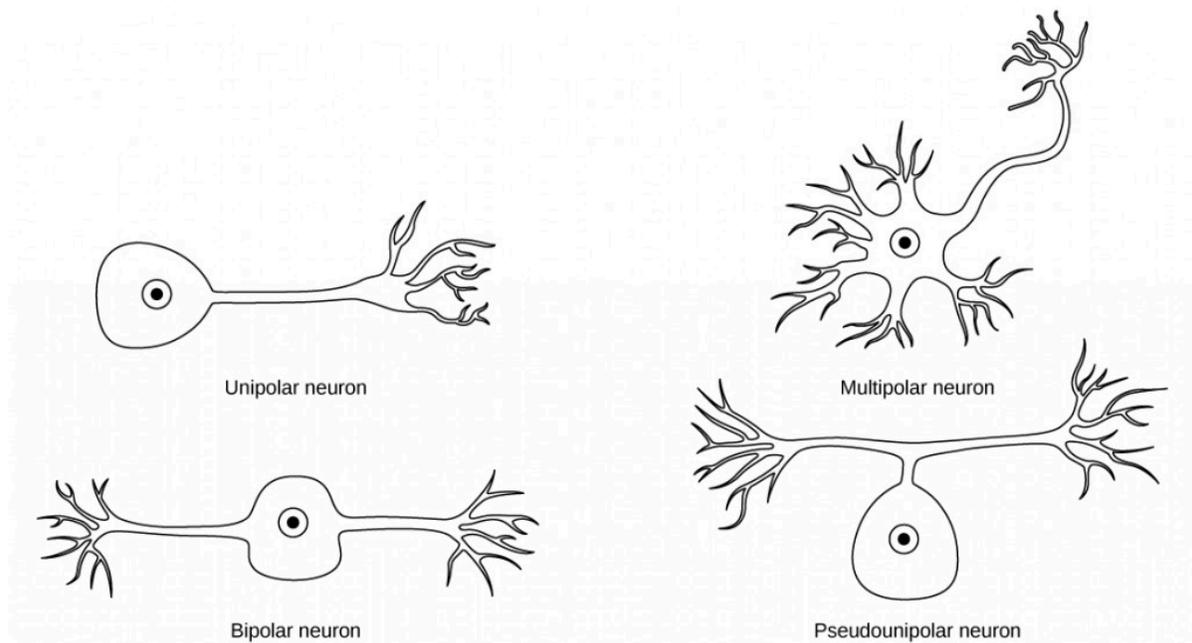
Figure 7.7. Neurons are broadly divided into four main types based on the number and placement of axons: (1) unipolar, (2) bipolar, (3) multipolar, and (4) pseudounipolar.

Neurogenesis

At one time, scientists believed that people were born with all the neurons they would ever have. Research performed during the last few decades indicates that neurogenesis, the birth of new neurons, continues into adulthood. Neurogenesis was first discovered in songbirds that produce new neurons while learning songs. For mammals, new neurons also play an important role in learning: about 1000 new neurons develop in the hippocampus (a brain structure involved in learning and memory) each day. While most of the new neurons will die, researchers found that an increase in the number of surviving new neurons in the hippocampus correlated with how well rats learned a new task. Interestingly, both exercise and some antidepressant medications also promote neurogenesis in the hippocampus. Stress has the opposite effect. While neurogenesis is quite limited compared to regeneration in other tissues, research in this area may lead to new treatments for disorders such as Alzheimer's, stroke, and epilepsy.

How do scientists identify new neurons? A researcher can inject a compound called bromodeoxyuridine (BrdU) into the brain of an animal. While all cells will be exposed to BrdU, BrdU will only be incorporated into the DNA of newly generated cells that are in S phase. A technique called immunohistochemistry can be used to attach a fluorescent label to the incorporated BrdU, and a researcher can use fluorescent microscopy to visualize the presence of BrdU, and thus new

neurons, in brain tissue. [Figure 7.8](#) is a micrograph which shows fluorescently labeled neurons in the hippocampus of a rat.
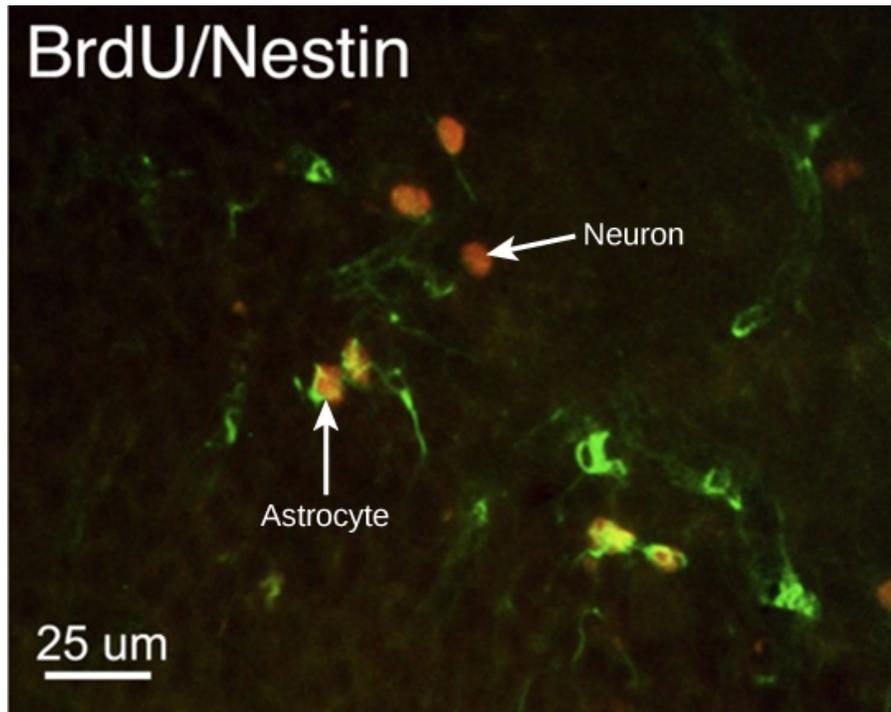
Figure 7.8. This micrograph shows fluorescently labeled new neurons in a rat hippocampus. Cells that are actively dividing have bromodoxyuridine (BrdU) incorporated into their DNA and are labeled in red. Cells that express glial fibrillary acidic protein (GFAP) are labeled in green. Astrocytes, but not neurons, express GFAP. Thus, cells that are labeled both red and green are actively dividing astrocytes, whereas cells labeled red only are actively dividing neurons. (credit: modification of work by Dr. Maryam Faiz, et. al., University of Barcelona; scale-bar data from Matt Russell)

[This site](#) contains more information about neurogenesis, including an interactive laboratory simulation and a video that explains how BrdU labels new cells.

Nerve impulse transmission within a neuron

For the nervous system to function, neurons must be able to send and receive signals. These signals

are possible because each neuron has a charged cellular membrane (a voltage difference between the inside and the outside), and the charge of this membrane can change in response to neurotransmitter molecules released from other neurons and environmental stimuli. To understand how neurons communicate, one must first understand the basis of the baseline or 'resting' membrane charge.

Neuronal charged membranes

The lipid bilayer membrane that surrounds a neuron is impermeable to charged molecules or ions. To enter or exit the neuron, ions must pass through special proteins called ion channels that span the membrane. Ion channels have different configurations: open, closed, and inactive, as illustrated in Figure 7.9. Some ion channels need to be activated in order to open and allow ions to pass into or out of the cell. These ion channels are sensitive to the environment and can change their shape accordingly. Ion channels that change their structure in response to voltage changes are called voltage-gated ion channels. Voltage-gated ion channels regulate the relative concentrations of different ions inside and outside the cell. The difference in total charge between the inside and outside of the cell is called the **membrane potential**.

Voltage-gated Na+ Channels

Outside cell

Cytoplasm

Closed At the resting potential, the channel is closed.

Open In response to a nerve impulse, the gate opens and Na+ enters the cell.

Inactivated For a brief period following activation, the channel does not open in response to a new signal.
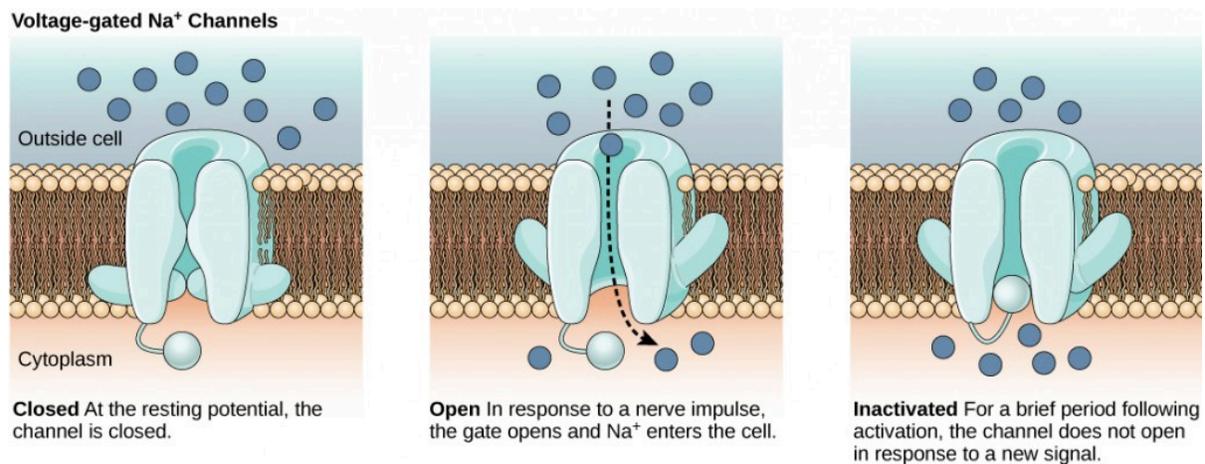
Figure 7.9. Voltage-gated ion channels open in response to changes in membrane voltage. After activation, they become inactivated for a brief period and will no longer open in response to a signal.

 This video discusses the basis of the resting membrane potential.

Resting membrane potential

A neuron at rest is negatively charged: the inside of a cell is approximately 70 millivolts more negative than the outside (–70 mV, note that this number varies by neuron type and by species). This voltage is called the resting membrane potential; it is caused by differences in the concentrations of ions inside and outside the cell. If the membrane were equally permeable to all ions, each type of ion would flow across the membrane and the system would reach equilibrium. Because ions cannot simply cross the membrane at will, there are different concentrations of several ions inside and outside the cell, as shown in Table 7.1. The difference in the number of positively charged potassium ions (K+) inside and outside the cell dominates the resting membrane potential ([Figure 7.10](#)). When the membrane is at rest, K+ions accumulate inside the cell due to a net movement with the concentration gradient. The negative resting membrane potential is created and maintained by increasing the concentration of cations outside the cell (in the extracellular fluid) relative to inside the cell (in the cytoplasm). The negative charge within the cell is created by the cell membrane being more permeable to potassium ion movement than sodium ion movement. In neurons, potassium ions are maintained at high concentrations within the cell while sodium ions are maintained at high concentrations outside of the cell. The cell possesses potassium and sodium leakage channels that allow the two cations to diffuse down their concentration gradient. However, the neurons have far more potassium leakage channels than sodium leakage channels. Therefore, potassium diffuses out of the cell at a much faster rate than sodium leaks in. Because more cations are leaving the cell than are entering, this causes the interior of the cell to be negatively charged relative to the outside of the cell. The actions of the sodium-potassium pump help to maintain the resting potential, once established. Recall that sodium potassium pumps bring two K+ ions into the cell while removing three Na+ ions per ATP consumed. As more cations are expelled from the cell than taken in, the inside of the cell remains negatively charged relative to the extracellular fluid. It should be noted that calcium ions (Cl–) tend to accumulate outside of the cell because they are repelled by negatively-charged proteins within the cytoplasm.

Table 7.1. The resting membrane potential is a result of different concentrations inside and outside the cell.

Ion Concentration Inside and Outside Neurons

Ion	Extracellular concentration (mM)	Intracellular concentration (mM)	Ratio outside,
Na+	145	12	12
K+	4	155	0.026
Cl–	120	4	30
Organic anions (A–)	–	100	

(a) Resting potential

0 mV

−70 mV   +40 mV

Extracellular fluid

Na⁺ Na⁺ K⁺ Na⁺ Na⁺ Na⁺ Na⁺

Na⁺ Na⁺ Na⁺ K⁺ channel Na⁺ Na⁺ Na⁺ Na⁺ channel

Cystoplasm

K⁺ K⁺ K⁺ K⁺ K⁺ K⁺ Na⁺ K⁺

K⁺ K⁺ K⁺ K⁺ K⁺ ATP ADP K⁺

Na⁺/K⁺ transporter K⁺ Na⁺ K⁺ K⁺

At the resting potential, all voltage-gated Na⁺ channels and most voltage-gated K⁺ channels are closed. The Na⁺/K⁺ transporter pumps K⁺ ions into the cell and Na⁺ ions out.

(b) Depolarization

0 mV

−70 mV   +40 mV

Na⁺ Na⁺ Na⁺ Na⁺ Na⁺

Na⁺ Na⁺ K⁺channel Na⁺ K⁺ Na⁺ Na⁺ Na⁺ Na⁺ channel Na⁺

K⁺ K⁺ K⁺ K⁺

K⁺ K⁺ K⁺ K⁺ Na⁺ Na⁺ K⁺

K⁺ Na⁺/K⁺ transporter K⁺ Na⁺ K⁺ K⁺ K⁺

In response to a depolarization, some Na⁺ channels open, allowing Na⁺ ions to enter the cell. The membrane starts to depolarize (the charge across the membrane lessens). If the threshold of excitation is reached, all the Na⁺ channels open.

(c) Hyperpolarization

0 mV

−70 mV   +40 mV

Na⁺ K⁺ K⁺ K⁺

K⁺ K⁺ K⁺ K⁺channel K⁺ Na⁺ K⁺ Na⁺ channel

Na⁺ Na⁺ Na⁺ Na⁺ Na⁺

Na⁺ Na⁺ Na⁺ Na⁺ Na⁺ Na⁺

Na⁺ K⁺ Na⁺/K⁺ transporter Na⁺ Na⁺ Na⁺

At the peak action potential, Na⁺ channels close while K⁺ channels open. K⁺ leaves the cell, and the membrane eventually becomes hyperpolarized.

Figure 7.10. The (a) resting membrane potential is a result of different concentrations of Na+ and K+ ions inside and outside the cell. A nerve impulse causes Na+ to enter the cell, resulting in (b) depolarization. At the peak action potential, K+ channels open and the cell becomes (c) hyperpolarized.

Action potential

A neuron can receive input from other neurons and, if this input is strong enough, send the signal to downstream neurons. Transmission of a signal between neurons is generally carried by a chemical called a neurotransmitter. Transmission of a signal within a neuron (from dendrite to axon terminal)

is carried by a brief reversal of the resting membrane potential called an **action potential**. When neurotransmitter molecules bind to receptors located on a neuron's dendrites, ion channels open. At excitatory synapses, this opening allows positive ions to enter the neuron and results in **depolarization** of the membrane—a decrease in the difference in voltage between the inside and outside of the neuron. A stimulus from a sensory cell or another neuron depolarizes the target neuron to its threshold potential (-55 mV). Na+ channels in the axon hillock open, allowing positive ions to enter the cell ([Figure 7.10](#) and [Figure 7.11](#)). Once the sodium channels open, the neuron completely depolarizes to a membrane potential of about +40 mV. Action potentials are considered an "all-or-nothing" event, in that, once the threshold potential is reached, the neuron always completely depolarizes. Once depolarization is complete, the cell must now "reset" its membrane voltage back to the resting potential. To accomplish this, the Na+ channels close and cannot be opened. This begins the neuron's **refractory period**, in which it cannot produce another action potential because its sodium channels will not open. At the same time, voltage-gated K+channels open, allowing K+ to leave the cell. As K+ ions leave the cell, the membrane potential once again becomes negative. The diffusion of K+ out of the cell actually **hyperpolarizes** the cell, in that the membrane potential becomes more negative than the cell's normal resting potential. At this point, the sodium channels will return to their resting state, meaning they are ready to open again if the membrane potential again exceeds the threshold potential. Eventually, the extra K+ ions diffuse out of the cell through the potassium leakage channels, bringing the cell from its hyperpolarized state, back to its resting membrane potential.
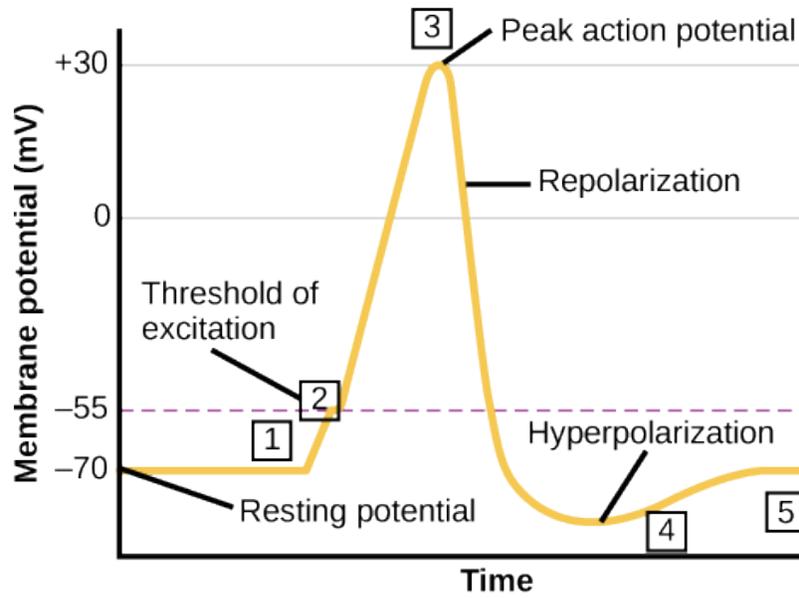
Figure 7.11. The formation of an action potential can be divided into five steps: (1) A stimulus from a sensory cell or another neuron causes the target cell to depolarize toward the threshold potential. (2) If the threshold of excitation is reached, all Na+ channels open and the membrane depolarizes. (3) At the peak action potential, K+ channels open and K+ begins to leave the cell. At the same time, Na+ channels close. (4) The membrane becomes hyperpolarized as K+ ions continue to leave the cell. The hyperpolarized membrane is in a refractory period and cannot fire. (5) The K+ channels close and the Na+/K+ transporter restores the resting potential.

In summary, an action potential is caused by movements of ions across the cell membrane as shown. Depolarization occurs when a stimulus makes the membrane permeable to Na^+ ions. Repolarization follows as the membrane again becomes impermeable to Na^+, and K^+ moves from high to low concentration. In the long term, active transport slowly maintains the concentration differences, but the cell may fire hundreds of times in rapid succession without seriously depleting them.

The separation of charge creates a potential difference of 70 to 90 mV across the cell membrane. While this is a small voltage, the resulting electric field ($E = V/d$) across the only 8-nm-thick membrane is immense (on the order of 11 MV/m!) and has fundamental effects on its structure and permeability. Now, if the exterior of a neuron is taken to be at 0 V, then the interior has a *resting potential* of about −90 mV. Such voltages are created across the membranes of almost all types of

animal cells but are largest in nerve and muscle cells. In fact, fully 25% of the energy used by cells goes toward creating and maintaining these potentials.

Electric currents along the cell membrane are created by any stimulus that changes the membrane's permeability. The membrane thus temporarily becomes permeable to Na^+, which then rushes in, driven both by diffusion and the Coulomb force. This inrush of Na^+ first neutralizes the inside membrane, or *depolarizes* it, and then makes it slightly positive. The depolarization causes the membrane to again become impermeable to Na^+, and the movement of K^+ quickly returns the cell to its resting potential, or *repolarizes* it. This sequence of events results in a voltage pulse called the *action potential* (Figure 7.12). Only small fractions of the ions move so that the cell can fire many hundreds of times without depleting the excess concentrations of Na^+ and K^+. Eventually, the cell must replenish these ions to maintain the concentration differences that create bioelectricity. This sodium-potassium pump is an example of *active transport*, wherein cell energy is used to move ions across membranes against diffusion gradients and the Coulomb force.
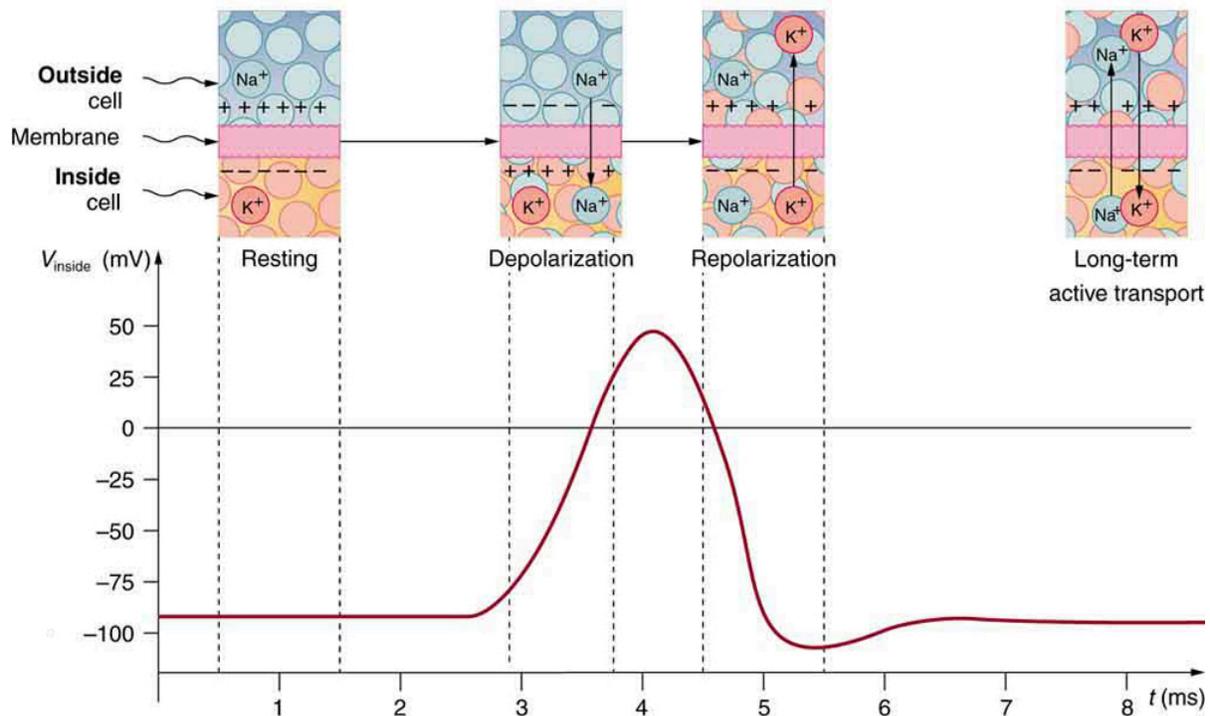
Figure 7.12. Action potential generation with illustrated ion movements at each step.

The action potential is a voltage pulse at one location on a cell membrane. How does it get transmitted along the cell membrane, and in particular down an axon, as a nerve impulse? The answer is that the changing voltage and electric fields affect the permeability of the adjacent cell membrane so that the same process takes place there. The adjacent membrane depolarizes, affecting the membrane further down, and so on, as illustrated in Figure 7.6. Thus the action potential stimulated at one location triggers a *nerve impulse* that moves slowly (about 1 m/s) along the cell membrane.

A nerve impulse is the propagation of an action potential along a cell membrane ([Figure 7.13](#)). A stimulus causes an action potential at one location, which changes the permeability of the adjacent membrane, causing an action potential there ([Figure 7.14](#)). This, in turn, affects the membrane further down, so that the action potential moves slowly (in electrical terms) along the cell membrane. Although the impulse is due to Na^+ and K^+ going across the membrane, it is equivalent to a wave of charge moving along the outside and inside of the membrane.
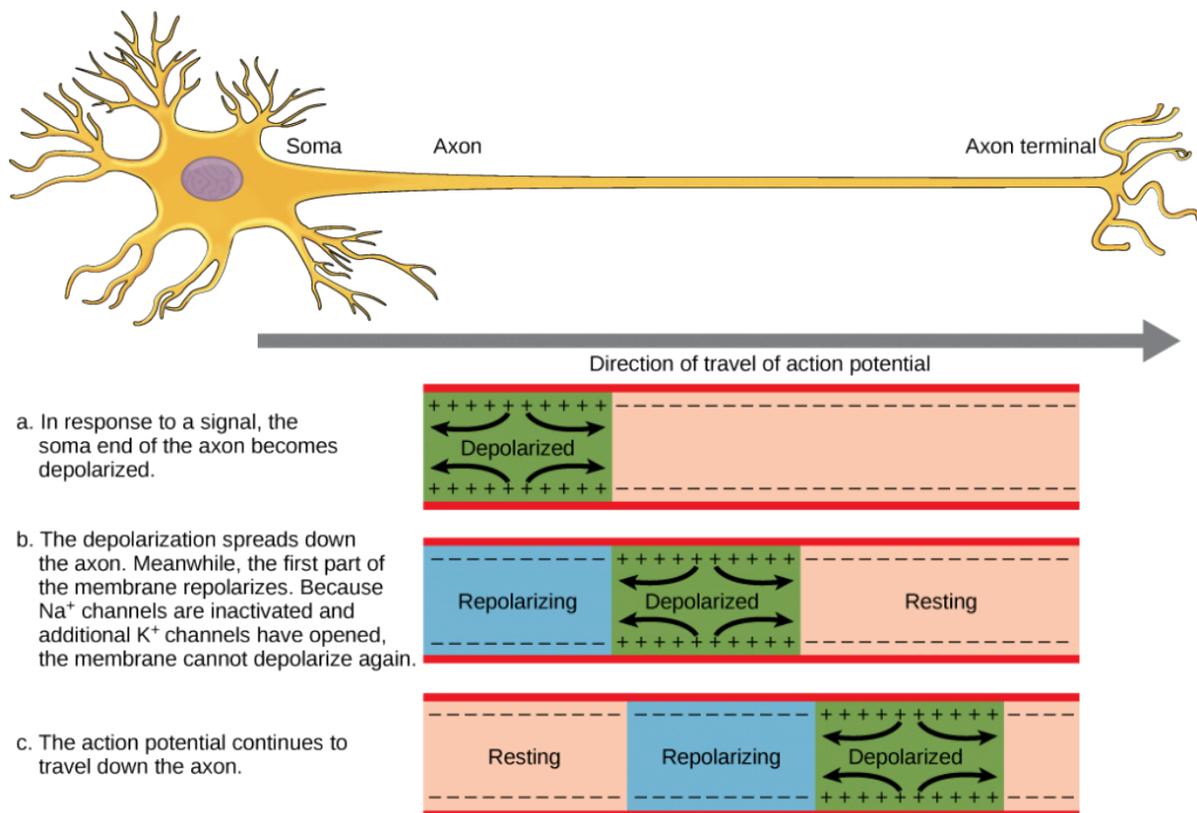
Figure 7.13. *The action potential is conducted down the axon as the axon membrane depolarizes, then repolarizes.*
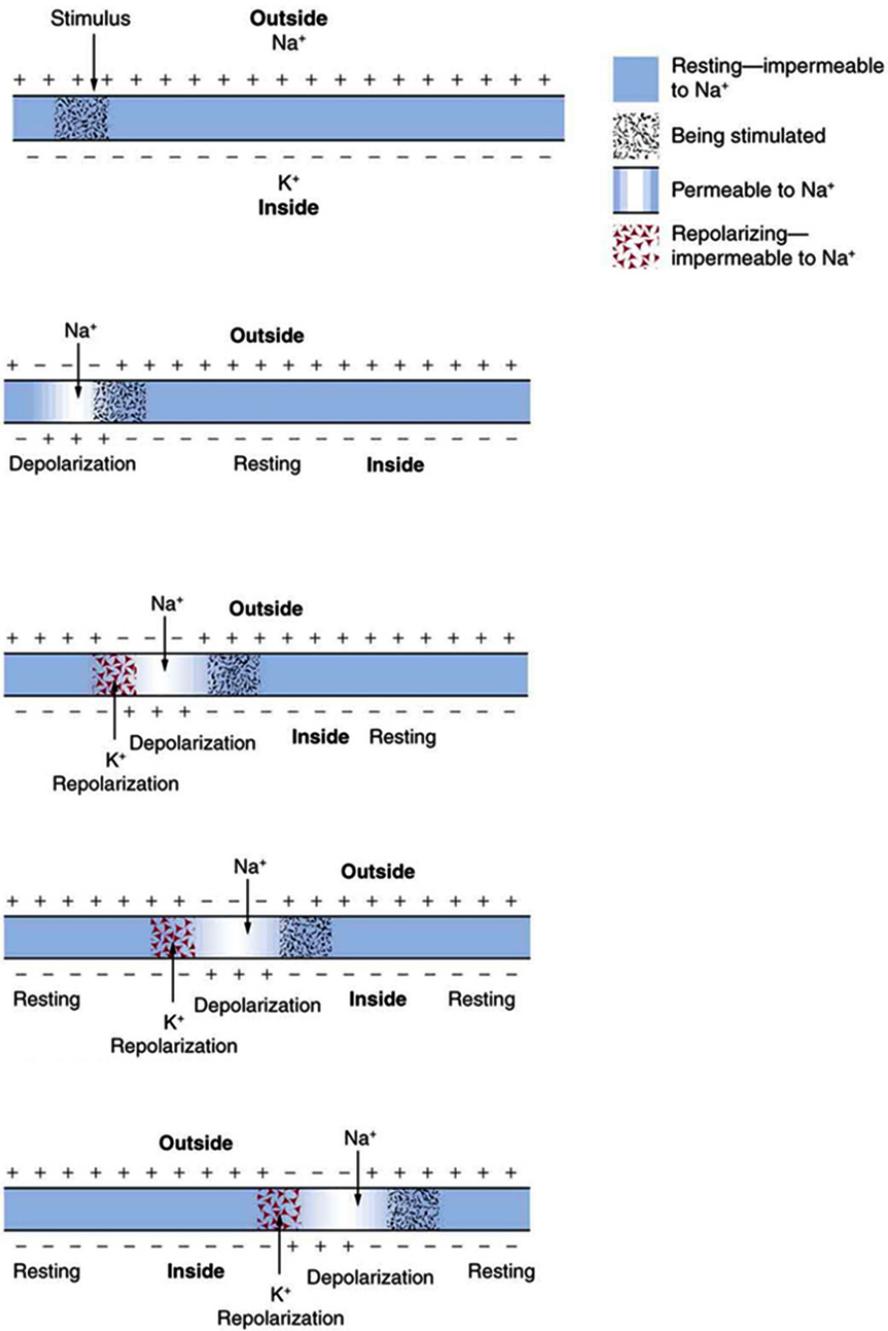
Figure 7.14. Propagation of an action potential.

This video presents an overview of an action potential.

Myelin and the propagation of the action potential

For an action potential to communicate information to another neuron, it must travel along the axon and reach the axon terminals where it can initiate neurotransmitter release. The speed of conduction of an action potential along an axon is influenced by both the diameter of the axon and the axon's resistance to current leak. Myelin acts as an insulator that prevents current from leaving the axon; this increases the speed of action potential conduction. In demyelinating diseases like multiple sclerosis, action potential conduction slows because the current leaks from previously insulated axon areas. The nodes of Ranvier, illustrated in Figure 7.15 are gaps in the myelin sheath along the axon. These unmyelinated spaces are about one micrometer long and contain voltage-gated Na+ and K+ channels. The flow of ions through these channels, particularly the Na+ channels, regenerates the action potential over and over again along the axon. This 'jumping' of the action potential from one node to the next is called **saltatory conduction**. If nodes of Ranvier were not present along an axon, the action potential would propagate very slowly since Na+ and K+ channels would have to continuously regenerate action potentials at every point along the axon instead of at specific points. Nodes of Ranvier also save energy for the neuron since the channels only need to be present at the nodes and not along the entire axon.
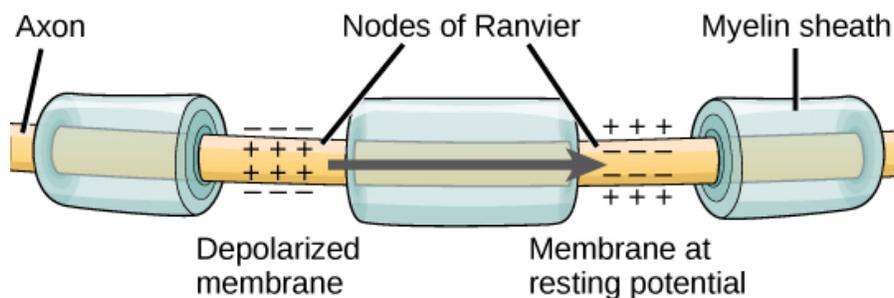
Figure 7.15. Nodes of Ranvier are gaps in myelin coverage along axons. Nodes contain voltage-gated K+ and Na+ channels. Action potentials travel down the axon by jumping from one node to the next.

Figure 7.16 shows an enlarged view of an axon having myelin sheaths characteristically separated by

unmyelinated gaps (called nodes of Ranvier). This arrangement gives the axon a number of interesting properties. Since myelin is an insulator, it prevents signals from jumping between adjacent nerves (crosstalk). Additionally, the myelinated regions transmit electrical signals at a very high speed, as an ordinary conductor or resistor would. There is no action potential in the myelinated regions so that no cell energy is used in them. There is an IR signal loss in the myelin, but the signal is regenerated in the gaps, where the voltage pulse triggers the action potential at full voltage. So a myelinated axon transmits a nerve impulse faster, with less energy consumption, and is better protected from cross talk than an unmyelinated one. Not all axons are myelinated so that crosstalk and slow signal transmission are a characteristic of the normal operation of these axons, another variable in the nervous system.

The degeneration or destruction of the myelin sheaths that surround the nerve fibers impairs signal transmission and can lead to numerous neurological effects. One of the most prominent of these diseases comes from the body's own immune system attacking the myelin in the central nervous system–multiple sclerosis. MS symptoms include fatigue, vision problems, weakness of arms and legs, loss of balance, and tingling or numbness in one's extremities (neuropathy). It is more apt to strike younger adults, especially females. Causes might come from infection, environmental or geographic effects, or genetics. At the moment there is no known cure for MS.

Most animal cells can fire or create their own action potential. Muscle cells contract when they fire and are often induced to do so by a nerve impulse. In fact, nerve and muscle cells are physiologically similar, and there are even hybrid cells, such as in the heart, that have characteristics of both nerves and muscles. Some animals, like the infamous electric eel ([Figure 7.17](#)) use muscles ganged so that their voltages add in order to create a shock great enough to stun prey.

Propagation of a nerve impulse down a myelinated axon, from left to right. The signal travels very fast and without energy input in the myelinated regions, but it loses voltage. It is regenerated in the gaps. The signal moves faster than in unmyelinated axons and is insulated from signals in other nerves, limiting crosstalk.
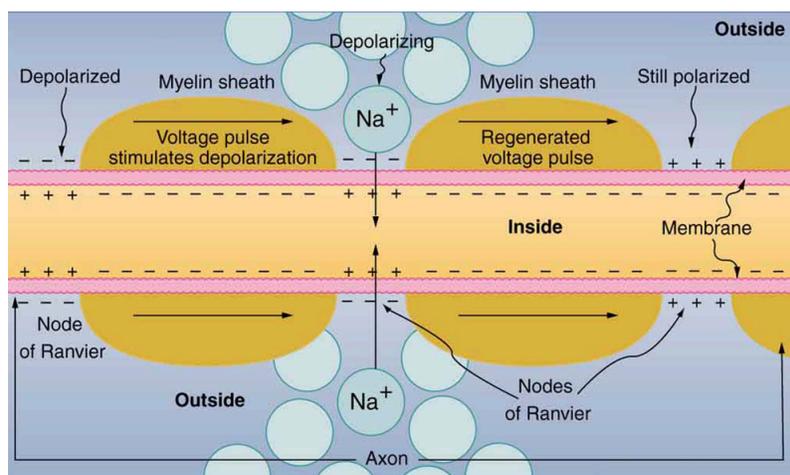

Figure 7.16. Propagation with myelin sheets present on a neuron.

Figure 7.17. An electric eel flexes its muscles to create a voltage that stuns prey. (credit: chrisbb, Flickr)

Question 7.5

Which of the following statements is false?
a. The soma is the cell body of a nerve cell.
b. Myelin sheath provides an insulating layer to the dendrites.
c. Axons carry the signal from the soma to the target.
d. Dendrites carry the signal to the soma.

Question 7.6

Neurons contain _____, which can receive signals from other neurons.
a. axons
b. mitochondria
c. dendrites
d. Golgi bodies

Question 7.7

A(n) _____ neuron has one axon and one dendrite extending directly from the cell body.
a. unipolar
b. bipolar
c. multipolar
d. pseudounipolar

Question 7.8

Glia that provide myelin for neurons in the brain are called _____.
a. Schwann cells
b. oligodendrocytes
c. microglia
d. astrocytes

Question 7.9

How are neurons similar to other cells? How are they unique?

Question 7.10

Compare and contrast resting, graded and action potential? In your answer, make sure you have included channels and voltage reference as well as relevant structures of the neurons. Once you have come up with an answer, give it to another student to review. Based on the review by your peer, is there something you need to work on in terms of your understanding of the resting, graded and action potential.

Question 7.11

Multiple sclerosis causes demyelination of axons in the brain and spinal cord. Why is this problematic?

7.3 – Synapses

7.3. Explain how sample signals cross the synapse.

Synaptic transmission

The synapse or "gap" is the place where information is transmitted from one neuron to another. Synapses usually form between axon terminals and dendritic spines, but this is not universally true. There are also axon-to-axon, dendrite-to-dendrite, and axon-to-cell body synapses. The neuron transmitting the signal is called the presynaptic neuron, and the neuron receiving the signal is called the postsynaptic neuron. Note that these designations are relative to a particular synapse—most neurons are both presynaptic and postsynaptic. There are two types of synapses: chemical and electrical.

Chemical synapse

When an action potential reaches the axon terminal it depolarizes the membrane and opens voltage-gated $Na+$ channels. $Na+$ ions enter the cell, further depolarizing the presynaptic membrane. This depolarization causes voltage-gated $Ca2+$ channels to open. Calcium ions entering the cell initiate a signaling cascade that causes small membrane-bound vesicles, called synaptic vesicles, containing neurotransmitter molecules to fuse with the presynaptic membrane. Synaptic vesicles are shown in Figure 7.18, which is an image from a scanning electron microscope.
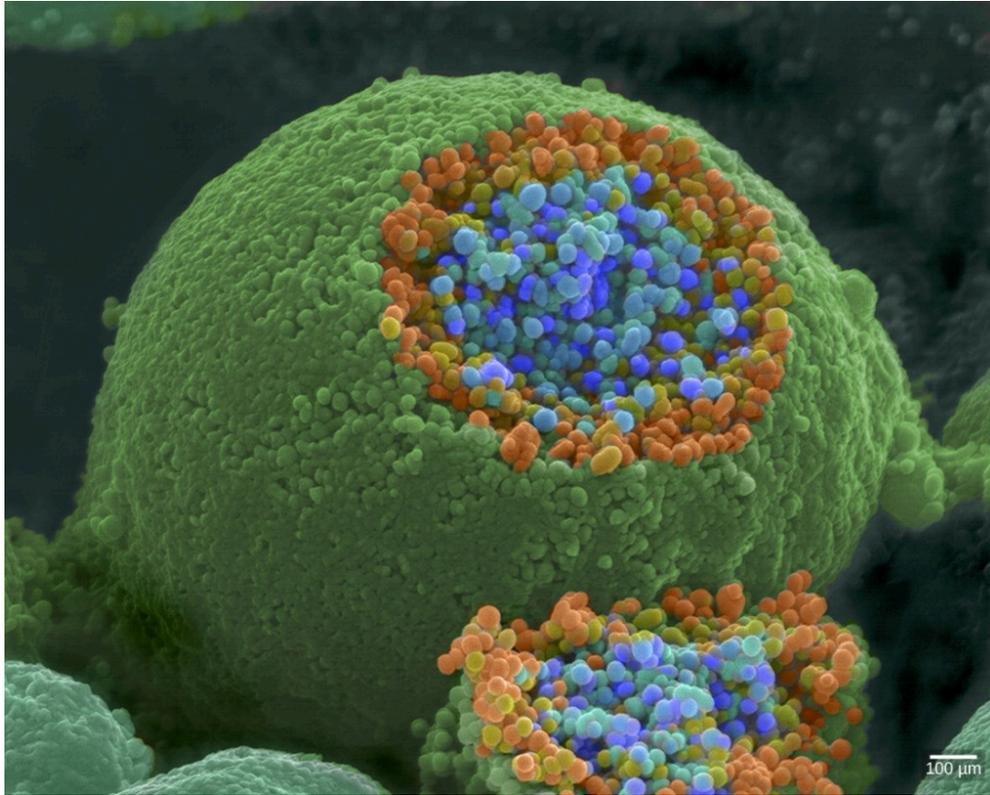
Figure 7.18. This pseudocolored image taken with a scanning electron microscope shows an axon terminal that was broken open to reveal synaptic vesicles (blue and orange) inside the neuron. (credit: modification of work by Tina Carvalho, NIH-NIGMS; scale-bar data from Matt Russell)

Fusion of a vesicle with the presynaptic membrane causes the neurotransmitter to be released into the synaptic cleft, the extracellular space between the presynaptic and postsynaptic membranes, as illustrated in Figure 7.19. The neurotransmitter diffuses across the synaptic cleft and binds to receptor proteins on the postsynaptic membrane.
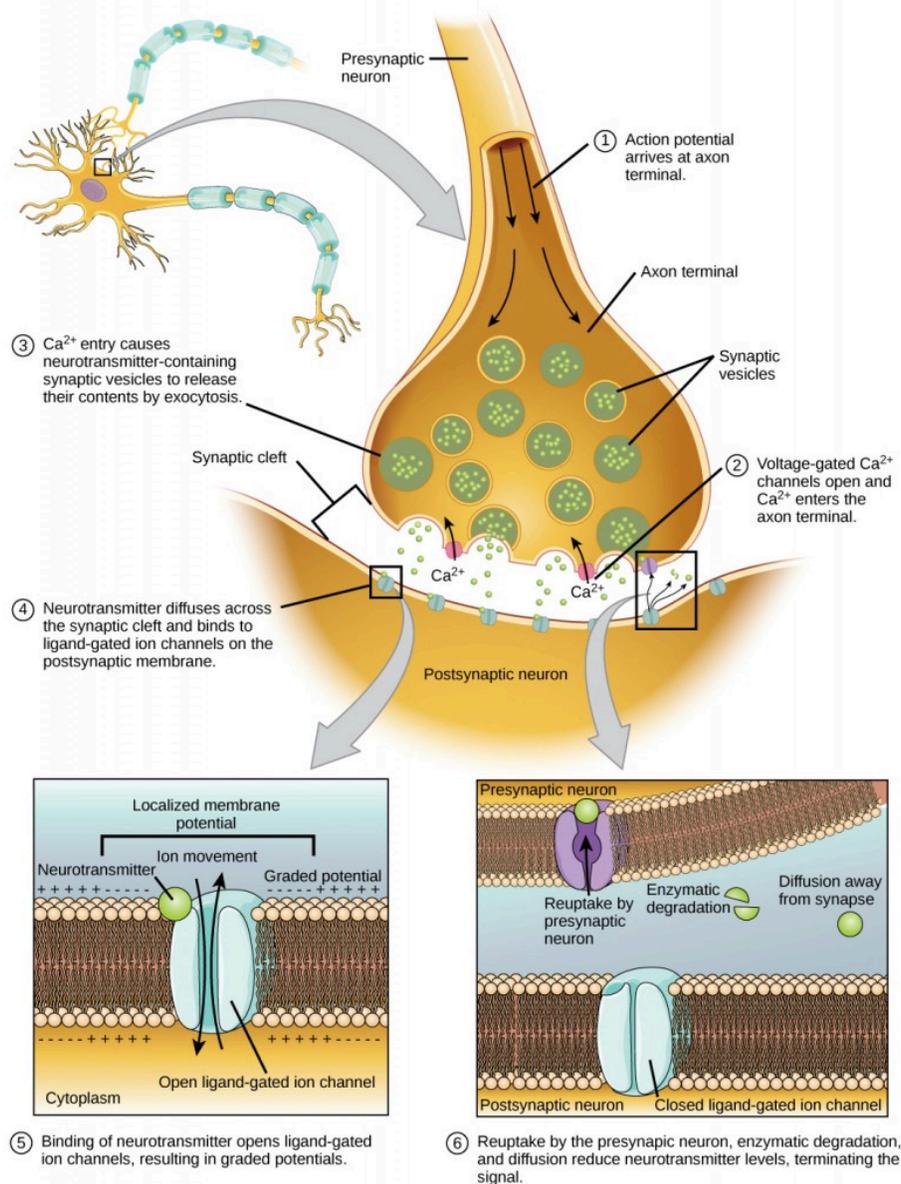
Figure 7.19. Communication at chemical synapses requires the release of neurotransmitters. When the presynaptic membrane is depolarized, voltage-gated Ca2+ channels open and allow Ca2+ to enter the cell. The calcium entry causes synaptic vesicles to fuse with the membrane and release neurotransmitter molecules into the synaptic cleft. The neurotransmitter diffuses across the synaptic cleft and binds to ligand-gated ion channels in the postsynaptic membrane, resulting in a localized depolarization or hyperpolarization of the postsynaptic neuron.

*The binding of a specific neurotransmitter causes particular ion channels, in this case, ligand-gated channels, on the postsynaptic membrane to open. Neurotransmitters can either have excitatory or inhibitory effects on the postsynaptic membrane, as detailed in Table 7.2. For example, when acetylcholine is released at the synapse between a nerve and muscle (called the neuromuscular junction) by a presynaptic neuron, it causes postsynaptic Na+ channels to open. Na+ enters the postsynaptic cell and causes the postsynaptic membrane to depolarize. This depolarization is called an **excitatory postsynaptic potential (EPSP)** and makes the postsynaptic neuron more likely*

*to fire an action potential. **The release** of neurotransmitter at inhibitory synapses causes **inhibitory postsynaptic potentials (IPSPs)**, a hyperpolarization of the presynaptic membrane. For example, when the neurotransmitter GABA (gamma-aminobutyric acid) is released from a presynaptic neuron, it binds to and opens Cl– channels. Cl– ions enter the cell and **hyperpolarizes** the membrane, making the neuron less likely to fire an action potential.*

Once neurotransmission has occurred, the neurotransmitter must be removed from the synaptic cleft so that the postsynaptic membrane can "reset" and be ready to receive another signal. This can be accomplished in three ways: the neurotransmitter can diffuse away from the synaptic cleft, it can be degraded by enzymes in the synaptic cleft, or it can be recycled (sometimes called reuptake) by the presynaptic neuron. Several drugs act at this step of neurotransmission. For example, some drugs that are given to Alzheimer's patients work by inhibiting acetylcholinesterase, the enzyme that degrades acetylcholine. This inhibition of the enzyme essentially increases neurotransmission at synapses that release acetylcholine. Once released, the acetylcholine stays in the cleft and can continually bind and unbind to postsynaptic receptors.

Table 7.2.

Neurotransmitter Function and Location

Neurotransmitter	Example	Location
Acetylcholine	–	CNS and/or PNS
Biogenic amine	Dopamine, serotonin, norepinephrine	CNS and/or PNS
Amino acid	Glycine, glutamate, aspartate, gamma-aminobutyric acid	CNS
Neuropeptide	Substance P, endorphins	CNS and/or PNS

Question 7.12

Design and/or act out events at the synapse! Here are some hints to help you design your review activity:
– You will need to think about how many peers you need to participate.
– Who will be which structure?
– How will your peers and you move with respect to each other?
– What is the overall goal of each movement you design (what is the function of each structure)?

Electrical synapse

While electrical synapses are fewer in number than chemical synapses, they are found in all nervous systems and play important and unique roles. The mode of neurotransmission in electrical synapses is quite different from that in chemical synapses. In an electrical synapse, the presynaptic and

postsynaptic membranes are very close together and are actually physically connected by channel proteins forming gap junctions. Gap junctions allow current to pass directly from one cell to the next. In addition to the ions that carry this current, other molecules, such as ATP, can diffuse through the large gap junction pores.

There are key differences between chemical and electrical synapses. Because chemical synapses depend on the release of neurotransmitter molecules from synaptic vesicles to pass on their signal, there is an approximately one-millisecond delay between when the axon potential reaches the presynaptic terminal and when the neurotransmitter leads to opening of postsynaptic ion channels. Additionally, this signaling is unidirectional. Signaling in electrical synapses, in contrast, is virtually instantaneous (which is important for synapses involved in key reflexes), and some electrical synapses are bidirectional. Electrical synapses are also more reliable as they are less likely to be blocked, and they are important for synchronizing the electrical activity of a group of neurons. For example, electrical synapses in the thalamus are thought to regulate slow-wave sleep, and disruption of these synapses can cause seizures.

Signal summation

Sometimes a single EPSP is strong enough to induce an action potential in the postsynaptic neuron, but often multiple presynaptic inputs must create EPSPs around the same time for the postsynaptic neuron to be sufficiently depolarized to fire an action potential. This process is called summation and occurs at the axon hillock, as illustrated in [Figure 7.20](#). Additionally, one neuron often has inputs from many presynaptic neurons—some excitatory and some inhibitory—so IPSPs can cancel out EPSPs and vice versa. It is the net change in postsynaptic membrane voltage that determines whether the postsynaptic cell has reached its threshold of excitation needed to fire an action potential. Together, synaptic summation and the threshold for excitation act as a filter so that random "noise" in the system is not transmitted as important information.
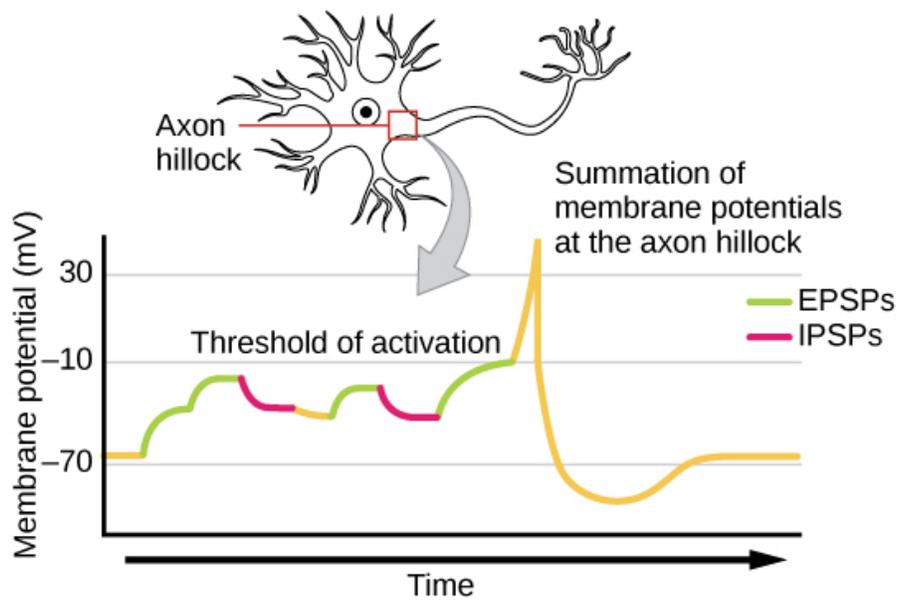
Figure 7.20. A single neuron can receive both excitatory and inhibitory inputs from multiple neurons, resulting in local membrane depolarization (EPSP input) and hyperpolarization (IPSP input). All these inputs are added together at the axon hillock. If the EPSPs are strong enough to overcome the IPSPs and reach the threshold of excitation, the neuron will fire.

Synaptic plasticity

Synapses are not static structures. They can be weakened or strengthened. They can be broken, and new synapses can be made. Synaptic plasticity allows for these changes, which are all needed for a functioning nervous system. In fact, synaptic plasticity is the basis of learning and memory. Two processes in particular, long-term potentiation (LTP) and long-term depression (LTD) are important forms of synaptic plasticity that occur in synapses in the hippocampus, a brain region that is involved in storing memories.

Long-term potentiation (LTP)

Long-term potentiation (LTP) is a persistent strengthening of a synaptic connection. LTP is based on the Hebbian principle: cells that fire together wire together. There are various mechanisms, none fully understood, behind the synaptic strengthening seen with LTP. One known mechanism involves a type of postsynaptic glutamate receptor, called NMDA (N-Methyl-D-aspartate) receptors, shown in Figure 7.21. These receptors are normally blocked by magnesium ions; however, when the postsynaptic neuron is depolarized by multiple presynaptic inputs in quick succession (either from

one neuron or multiple neurons), the magnesium ions are forced out allowing Ca ions to pass into the postsynaptic cell. Next, Ca2+ions entering the cell initiate a signaling cascade that causes a different type of glutamate receptor, called AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) receptors, to be inserted into the postsynaptic membrane, since activated AMPA receptors allow positive ions to enter the cell. So, the next time glutamate is released from the presynaptic membrane, it will have a larger excitatory effect (EPSP) on the postsynaptic cell because the binding of glutamate to these AMPA receptors will allow more positive ions into the cell. The insertion of additional AMPA receptors strengthens the synapse and means that the postsynaptic neuron is more likely to fire in response to presynaptic neurotransmitter release. Some drugs of abuse co-opt the LTP pathway, and this synaptic strengthening can lead to addiction.

Long-term depression (LTD)

Long-term depression (LTD) is essentially the reverse of LTP: it is a long-term weakening of a synaptic connection. One mechanism known to cause LTD also involves AMPA receptors. In this situation, calcium that enters through NMDA receptors initiates a different signaling cascade, which results in the removal of AMPA receptors from the postsynaptic membrane, as illustrated in [Figure 7.21](). The decrease in AMPA receptors in the membrane makes the postsynaptic neuron less responsive to glutamate released from the presynaptic neuron. While it may seem counterintuitive, LTD may be just as important for learning and memory as LTP. The weakening and pruning of unused synapses allow for unimportant connections to be lost and makes the synapses that have undergone LTP that much stronger by comparison.

The NMDA receptor is
activated by glutamate
binding, but only after
depolarization removes
inhibitory Mg^{2+}. Once the
Mg^{2+} is removed, Ca^{2+} can
enter the cell.

Some AMPA receptors are
present in the membrane
initially. In response to an
increase in intracellular Ca^{2+},
more are inserted.

Low-frequency stimulation
results in a different Ca^{2+}
-signaling cascade. AMPA
receptor is removed from
the membrane, and as a
result, the nerve cell
becomes less responsive
to glutamate.

Figure 7.21. Calcium entry through postsynaptic NMDA receptors can initiate two different forms of synaptic plasticity: long-term potentiation (LTP) and long-term depression (LTD). LTP arises when a single synapse is repeatedly stimulated. This stimulation causes a calcium- and CaMKII-dependent cellular cascade, which results in the insertion of more AMPA receptors into the postsynaptic membrane. The next time glutamate is released from the presynaptic cell, it will bind to both NMDA and the newly inserted AMPA receptors, thus depolarizing the membrane more efficiently. LTD occurs when few glutamate molecules bind to NMDA receptors at a synapse (due to a low firing rate of the presynaptic neuron). The calcium that does flow through NMDA receptors initiates a different calcineurin and protein phosphatase 1-dependent cascade, which results in the endocytosis of AMPA receptors. This makes the postsynaptic neuron less responsive to glutamate released from the presynaptic neuron.

Question 7.13

For a neuron to fire an action potential, its membrane must reach _____.
a. hyperpolarization
b. the threshold of excitation
c. the refractory period
d. inhibitory postsynaptic potential

Question 7.14

After an action potential, the opening of additional voltage-gated _____ channels and the inactivation of sodium channels, cause the membrane to return to its resting membrane potential.
a. sodium
b. potassium
c. calcium
d. chloride

Question 7.15

What is the term for protein channels that connect two neurons at an electrical synapse?
a. synaptic vesicles
b. voltage-gated ion channels
c. gap junction protein
d. sodium-potassium exchange pumps

Question 7.16

How does myelin aid propagation of an action potential along an axon? How do the nodes of Ranvier help this process?

Question 7.17

What are the main steps in chemical neurotransmission?

7.4 – Neuroglia

7.4. Describe the roles of different neuroglia.

Glia

While glia are often thought of as the supporting cast of the nervous system, the number of glial cells in the brain actually outnumbers the number of neurons by a factor of ten. Neurons would be unable to function without the vital roles that are fulfilled by these glial cells. Glia guide developing neurons to their destinations, buffer ions and chemicals that would otherwise harm neurons and provide myelin sheaths around axons. Scientists have recently discovered that they also play a role in responding to nerve activity and modulating communication between nerve cells. When glia do not function properly, the result can be disastrous–most brain tumors are caused by mutations in glia.

Glia have different roled in both the central nervous system (CNS) and the peripheral nervous system (PNS). CNS is comprised of the brain and the spinal chord while PNS is the connection between the central nervous system and the rest of the body. The CNS is like the power plant of the nervous system. It creates the signals that control the functions of the body. The PNS is like the wires that go to individual houses. Without those "wires," the signals produced by the CNS could not control the body (and the CNS would not be able to receive sensory information from the body either). The PNS can be broken down into the autonomic nervous system, which controls bodily functions without conscious control, and the sensory-somatic nervous system, which transmits sensory information from the skin, muscles, and sensory organs to the CNS and sends motor commands from the CNS to the muscles.

Types of Glia

There are several different types of glia with different functions, two of which are shown in Figure 7.22. **Astrocytes**, shown in Figure 7.23a make contact with both capillaries and neurons in the central nervous system (CNS). They provide nutrients and other substances to neurons, regulate the

concentrations of ions and chemicals in the extracellular fluid, and provide structural support for synapses. Astrocytes also form the blood-brain barrier–a structure that blocks the entrance of toxic substances into the brain. Astrocytes, in particular, have been shown through calcium imaging experiments to become active in response to nerve activity, transmit calcium waves between astrocytes, and modulate the activity of surrounding synapses. **Satellite glia** provide nutrients and structural support for neurons in the PNS. **Microglia** scavenge and degrade dead cells and protect the brain from invading microorganisms. **Oligodendrocytes**, shown in [Figure 7.23b](#) form myelin sheaths around axons in the CNS. One axon can be myelinated by several oligodendrocytes, and one oligodendrocyte can provide myelin for multiple neurons. This is distinctive from the PNS where a single **Schwann cell** provides myelin for only one axon as the entire Schwann cell surrounds the axon. **Radial glia** serve as scaffolds for developing neurons as they migrate to their end destinations. **Ependymal** cells line fluid-filled ventricles of the brain and the central canal of the spinal cord. They are involved in the production of cerebrospinal fluid, which serves as a cushion for the brain, moves the fluid between the spinal cord and the brain, and is a component for the choroid plexus.
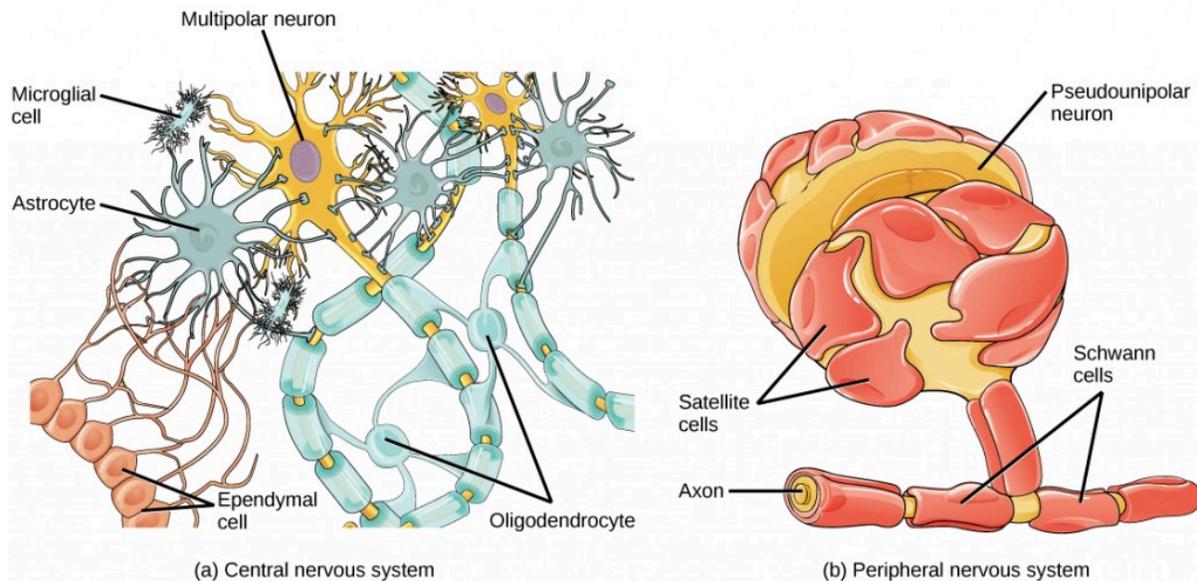
Figure 7.22. Glial cells support neurons and maintain their environment. Glial cells of the (a) central nervous system include oligodendrocytes, astrocytes, ependymal cells, and microglial cells. Oligodendrocytes form the myelin sheath around axons. Astrocytes provide nutrients to neurons, maintain their extracellular environment, and provide structural support. Microglia scavenge pathogens and dead cells. Ependymal cells produce cerebrospinal fluid that cushions the neurons. Glial cells of the (b) peripheral nervous system include Schwann cells, which form the myelin sheath, and satellite cells, which provide nutrients and structural support to neurons.
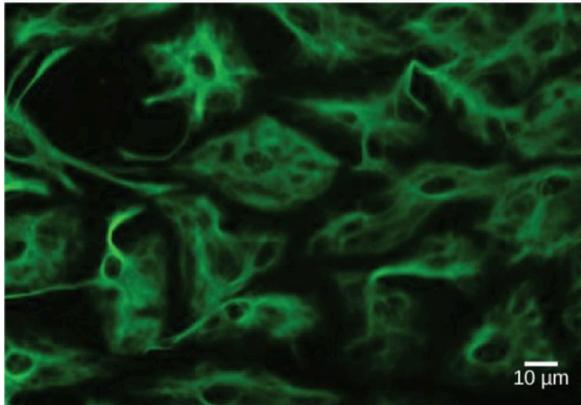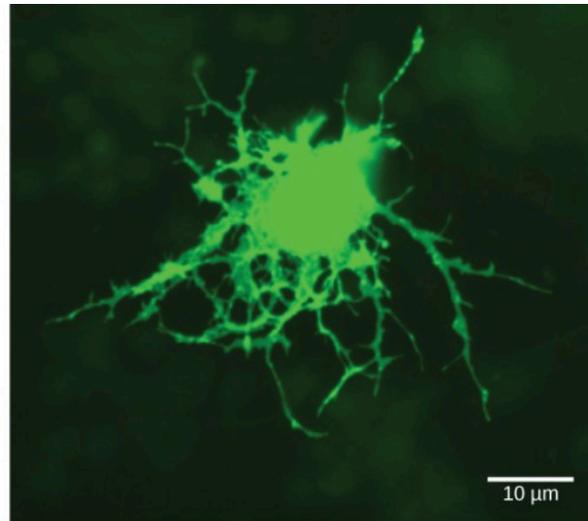
Figure 7.23. (a) Astrocytes and (b) oligodendrocytes are glial cells of the central nervous system. (credit a: modification of work by Uniformed Services University; credit b: modification of work by Jurjen Broeke; scale-bar data from Matt Russell)

Question 7.18

What is a functionally equivalent cell in CNS to a Schwann cell in PNS? What other types of glia exist in CNS and what is their function. Try drawing all the glia of the CNS this cell in addition to answering the above questions.

CHAPTER 8 – SENSORY SYSTEMS

8.1. Describe the general organization of the neurons within and outside of the central nervous system;

8.2. Summarize evolution of key brain features among animals; and,

8.3. Explain how olfactory, gustatory, equilibrium, auditory, and visual sensations are interpreted by animals with reference to signal processing and/or specific ion movements.

The content of this chapter was adapted from the *Concepts of Biology-1st Canadian Edition* open textbook by Charles Molnar and Jane Gair and *Anatomy and Physiology* open textbook.

8.1 – Neuronal organization

8.1. Describe the general organization of the neurons within and outside of the central nervous system.

Central nervous system

The central nervous system (CNS) is made up of the brain, a part of which is shown in [Figure 8.1](#) and spinal cord and is covered with three layers of protective coverings called **meninges** (from the Greek word for membrane). The outermost layer is the **dura mater** (Latin for "hard mother"). As the Latin suggests, the primary function of this thick layer is to protect the brain and spinal cord. The dura mater also contains vein-like structures that carry blood from the brain back to the heart. The middle layer is the web-like **arachnoid mater**. The last layer is the **pia mater** (Latin for "soft mother"), which directly contacts and covers the brain and spinal cord like plastic wrap. The space between the arachnoid and pia maters is filled with **cerebrospinal fluid (CSF)**. CSF is produced by a tissue called **choroid plexus** in fluid-filled compartments in the CNS called *ventricles*. The brain floats in CSF, which acts as a cushion and shock absorber and makes the brain neutrally buoyant. CSF also functions to circulate chemical substances throughout the brain and into the spinal cord.

The entire brain contains only about 8.5 tablespoons of CSF, but CSF is constantly produced in the ventricles. This creates a problem when a ventricle is blocked–the CSF builds up and creates swelling and the brain is pushed against the skull. This swelling condition is called hydrocephalus ("water head") and can cause seizures, cognitive problems, and even death if a shunt is not inserted to remove the fluid and pressure.
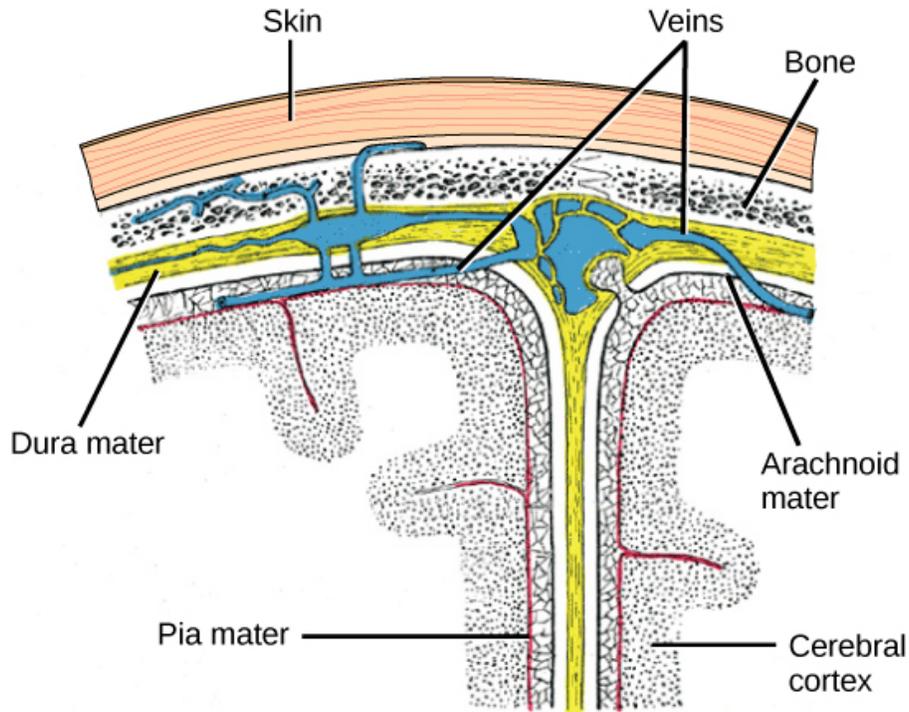
Figure 8.1. The cerebral cortex is covered by three layers of meninges: the dura, arachnoid, and pia maters. (credit: modification of work by Gray's Anatomy)

Autonomic nervous system

The autonomic nervous system serves as the relay between the CNS and the internal organs. It controls the lungs, the heart, smooth muscle, and exocrine and endocrine glands. The autonomic nervous system controls these organs largely without conscious control; it can continuously monitor the conditions of these different systems and implement changes as needed. Signaling to the target tissue usually involves two synapses: a preganglionic neuron (originating in the CNS) synapses to a neuron in a ganglion that, in turn, synapses on the target organ, as illustrated in Figure 8.2. There are two divisions of the autonomic nervous system that often have opposing effects: the sympathetic nervous system and the parasympathetic nervous system.
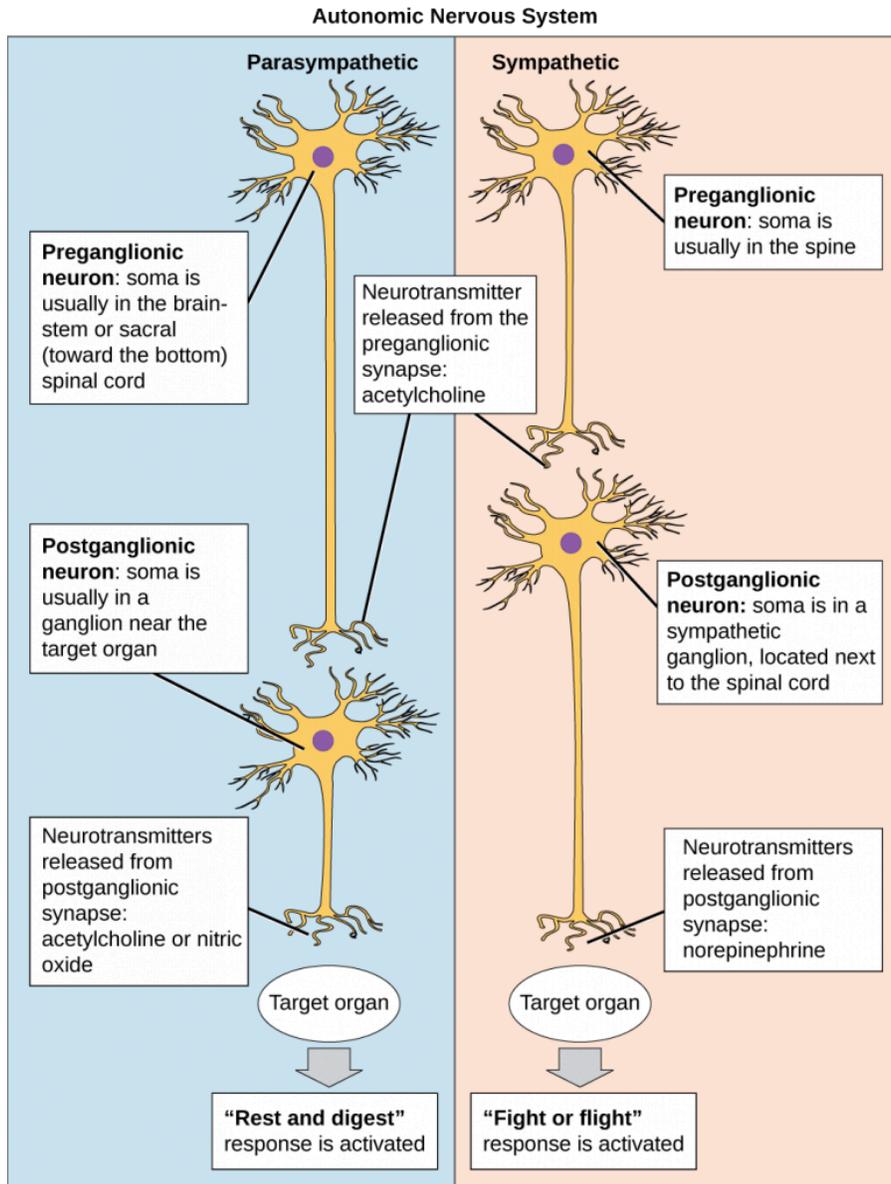
Figure 8.2. Autonomic nervous system.

Question 8.1

Which of the following statements is false?
a. The parasympathetic pathway is responsible for resting the body, while the sympathetic pathway is responsible for preparing for an emergency.
b. Most preganglionic neurons in the sympathetic pathway originate in the spinal cord.
c. Slowing of the heartbeat is a parasympathetic response.
d. Parasympathetic neurons are responsible for releasing norepinephrine on the target organ, while sympathetic neurons are responsible for releasing acetylcholine.

Sympathetic nervous system

The **sympathetic nervous system** is responsible for the "fight or flight" response that occurs when an animal encounters a dangerous situation. One way to remember this is to think of the surprise a person feels when encountering a snake ("snake" and "sympathetic" both begin with "s"). Examples of functions controlled by the sympathetic nervous system include an accelerated heart rate and inhibited digestion. These functions help prepare an organism's body for the physical strain required to escape a potentially dangerous situation or to fend off a predator.

Most preganglionic neurons in the sympathetic nervous system originate in the spinal cord, as illustrated in [Figure 8.3](#). The axons of these neurons release **acetylcholine** on postganglionic neurons within the sympathetic ganglia (the sympathetic ganglia form a chain that extends alongside the spinal cord). The acetylcholine activates the postganglionic neurons. Postganglionic neurons then release **norepinephrine** onto target organs. As anyone who has ever felt a rush before a big test, speech, or athletic event can attest, the effects of the sympathetic nervous system are quite pervasive. This is both because one preganglionic neuron synapses on multiple postganglionic neurons, amplifying the effect of the original synapse, and because the adrenal gland also releases norepinephrine (and the closely related hormone epinephrine) into the bloodstream. The physiological effects of this norepinephrine release include dilating the trachea and bronchi (making it easier for the animal to breathe), increasing heart rate, and moving blood from the skin to the heart, muscles, and brain (so the animal can think and run). The strength and speed of the sympathetic response help an organism avoid danger, and scientists have found evidence that it may also increase LTP–allowing the animal to remember the dangerous situation and avoid it in the future.
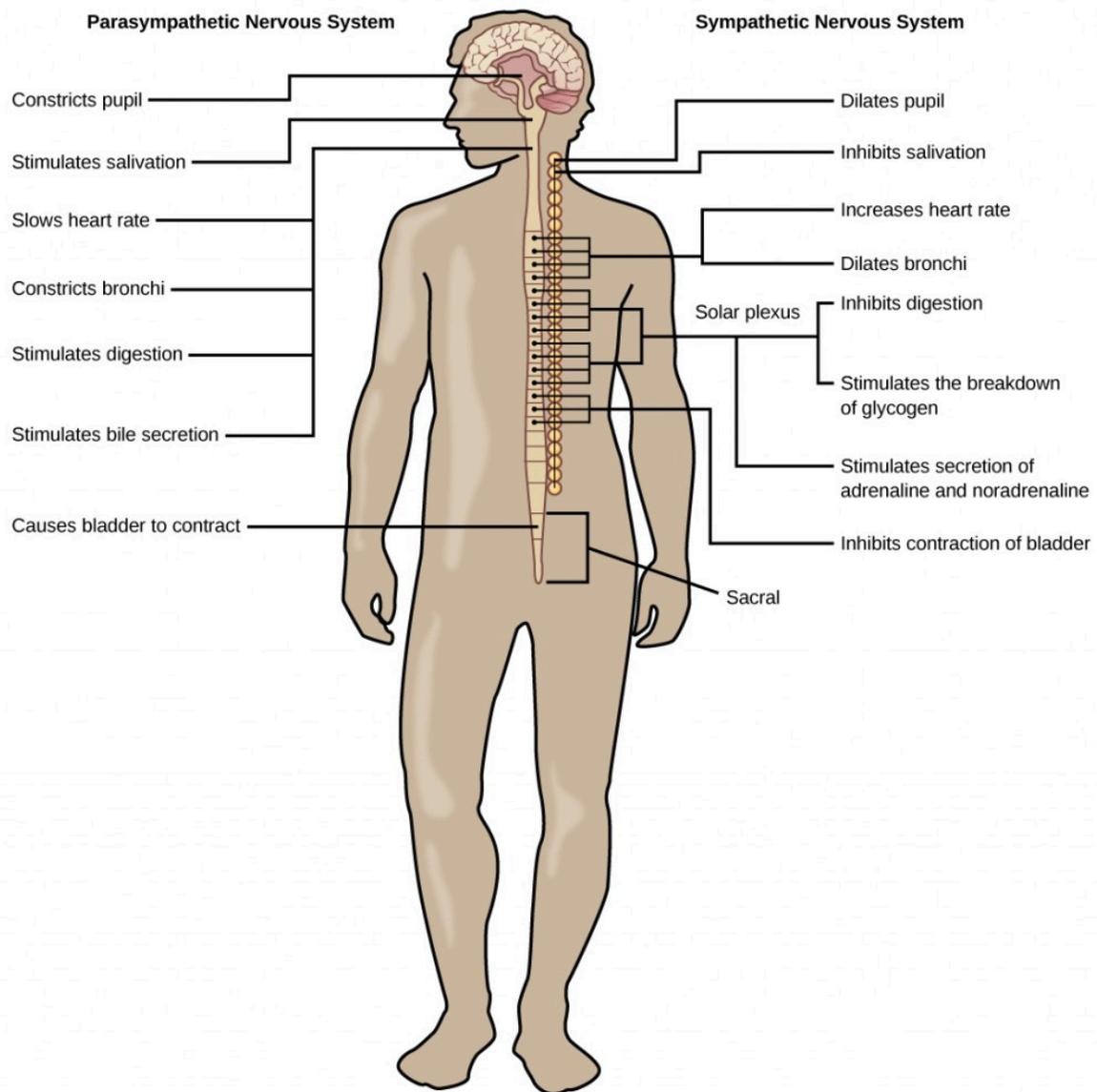
Figure 8.3. Sympathetic and parasympathetic nervous system in humans.

Parasympathetic nervous system

While the sympathetic nervous system is activated in stressful situations, the **parasympathetic nervous system** allows an animal to "rest and digest." One way to remember this is to think that during a restful situation like a picnic, the parasympathetic nervous system is in control ("picnic" and "parasympathetic" both start with "p"). Parasympathetic preganglionic neurons have cell bodies located in the brainstem and in the sacral (toward the bottom) spinal cord, as shown in Figure 8.3.

The axons of the preganglionic neurons release acetylcholine on the postganglionic neurons, which are generally located very near the target organs. Most postganglionic neurons release acetylcholine onto target organs, although some release nitric oxide.

The parasympathetic nervous system resets organ function after the sympathetic nervous system is activated (the common adrenaline dump you feel after a 'fight-or-flight' event). Effects of acetylcholine release on target organs include slowing of heart rate, lowered blood pressure, and stimulation of digestion.

Sensory-somatic nervous system

The sensory-somatic nervous system is made up of cranial and spinal nerves and contains both sensory and motor neurons. Sensory neurons transmit sensory information from the skin, skeletal muscle, and sensory organs to the CNS. Motor neurons transmit messages about the desired movement from the CNS to the muscles to make them contract. Without its sensory-somatic nervous system, an animal would be unable to process any information about its environment (what it sees, feels, hears, and so on) and could not control motor movements. Unlike the autonomic nervous system, which has two synapses between the CNS and the target organ, sensory and motor neurons have only one synapse—one ending of the neuron is at the organ and the other directly contacts a CNS neuron. Acetylcholine is the main neurotransmitter released at these synapses.

Humans have 12 **cranial nerves**, nerves that emerge from or enter the skull (cranium), as opposed to the spinal nerves, which emerge from the vertebral column. Each cranial nerve is accorded a name, which is detailed in [Figure 8.4](). Some cranial nerves transmit only sensory information. For example, the olfactory nerve transmits information about smells from the nose to the brainstem. Other cranial nerves transmit almost solely motor information. For example, the oculomotor nerve controls the opening and closing of the eyelid and some eye movements. Other cranial nerves contain a mix of sensory and motor fibers. For example, the glossopharyngeal nerve has a role in both taste (sensory) and swallowing (motor).

Spinal nerves transmit sensory and motor information between the spinal cord and the rest of the body. Each of the 31 spinal nerves (in humans) contains both sensory and motor axons. The sensory neuron cell bodies are grouped in structures called dorsal root ganglia and are shown in [Figure 8.5](). Each sensory neuron has one projection—with a sensory receptor ending in the skin, muscle, or sensory organs—and another that synapses with a neuron in the dorsal spinal cord. Motor neurons have cell bodies in the ventral gray matter of the spinal cord that project to muscle through the ventral root. These neurons are usually stimulated by interneurons within the spinal cord but are sometimes directly stimulated by sensory neurons.
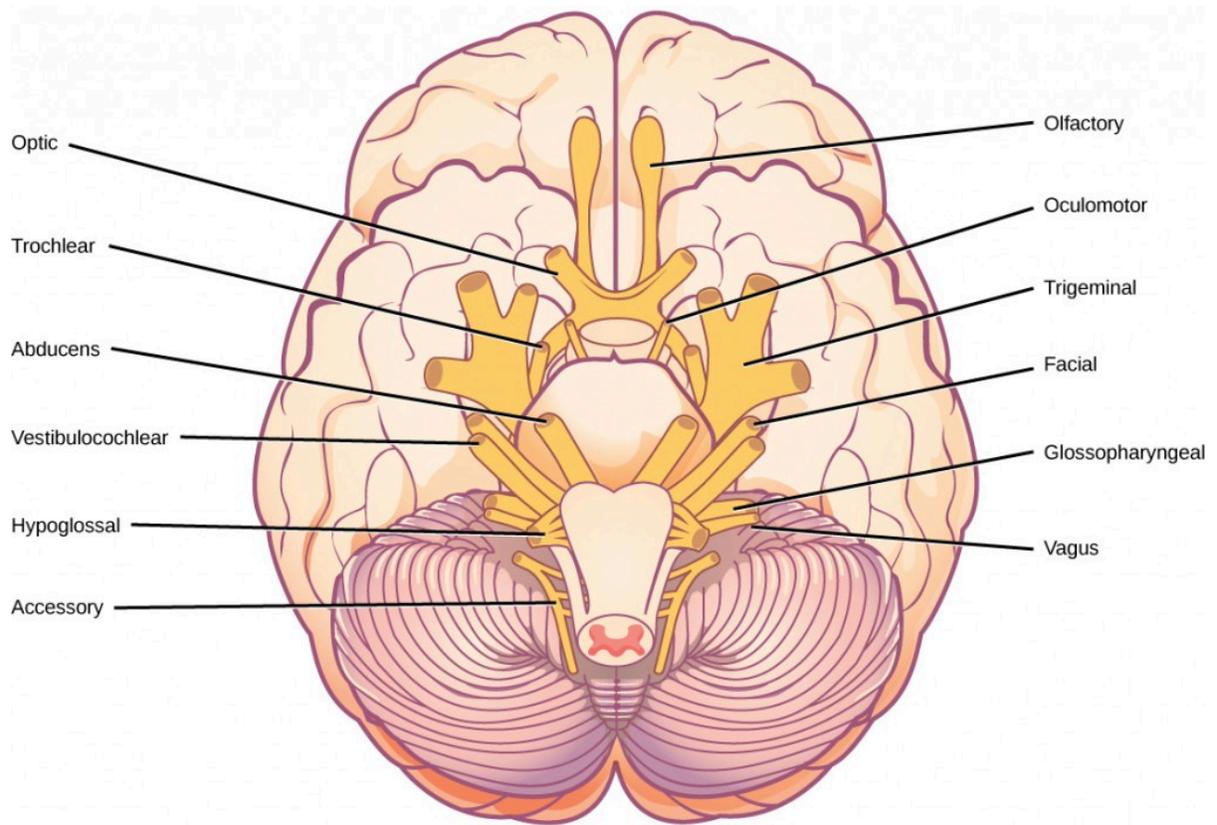
Figure 8.4. Cranial nerves in humans.

White matter

Gray matter

Dorsal root

Dorsal root ganglion

Sensory neuron soma

Motor neuron soma

Spinal nerve

Ventral root

Cross Section of Spinal Cord
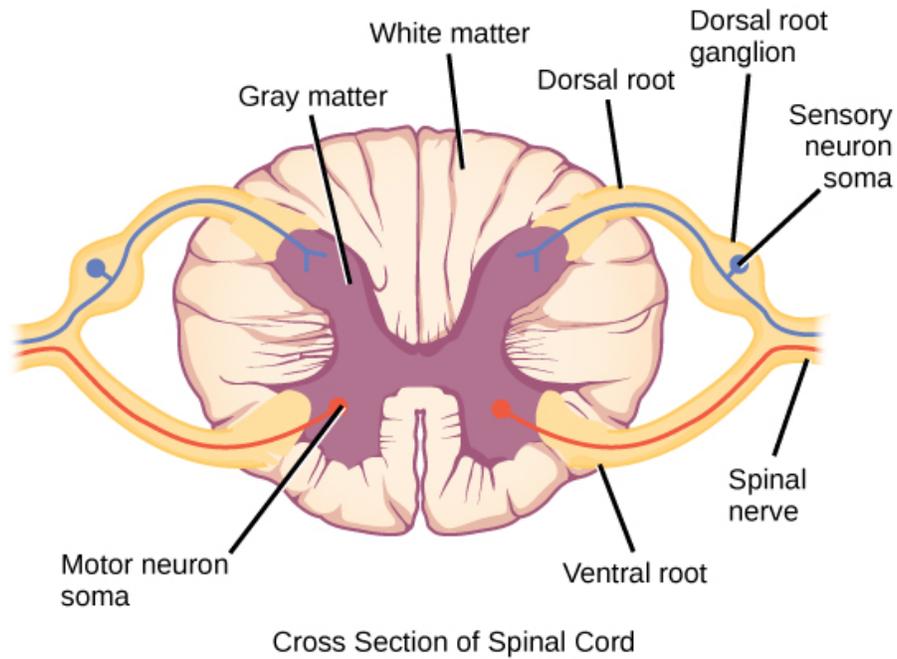
Figure 8.5. Spinal cord structures.

The peripheral nervous system contains both the autonomic and sensory-somatic nervous systems. The autonomic nervous system provides unconscious control over visceral functions and has two divisions: the sympathetic and parasympathetic nervous systems. The sympathetic nervous system is activated in stressful situations to prepare the animal for a "fight or flight" response. The parasympathetic nervous system is active during restful periods. The sensory-somatic nervous system is made of cranial and spinal nerves that transmit sensory information from skin and muscle to the CNS and motor commands from the CNS to the muscles.

?

Question 8.2

When you use a word nerve what does it refer to? What about neuron? How are these terms different and how are they the same in terms of their function.

8.2 – Brain evolution

8.2. Summarize evolution of key brain features among animals.

Organization of the central nervous system

When you're reading this book, your nervous system is performing several functions simultaneously. The visual system is processing what is seen on the page; the motor system controls the turn of the pages (or click of the mouse); the prefrontal cortex maintains attention. Even fundamental functions, like breathing and regulation of body temperature, are controlled by the nervous system. A nervous system is an organism's control center: it processes sensory information from outside (and inside) the body and controls all behaviors—from eating to sleeping to finding a mate.

Nervous systems throughout the animal kingdom vary in structure and complexity, as illustrated by the variety of animals shown in [Figure 8.6]. Some organisms, like sea sponges, lack a true nervous system. Others, like jellyfish, lack a true brain and instead have a system of separate but connected nerve cells (neurons) called a "nerve net." Echinoderms such as sea stars have nerve cells that are bundled into fibers called nerves. Flatworms of the phylum Platyhelminthes have both a central nervous system (CNS), made up of a small "brain" and two nerve cords and a peripheral nervous system (PNS) containing a system of nerves that extend throughout the body. The insect nervous system is more complex but also fairly decentralized. It contains a brain, ventral nerve cord, and ganglia (clusters of connected neurons). These ganglia can control movements and behaviors without input from the brain. Octopi may have the most complicated of invertebrate nervous systems—they have neurons that are organized in specialized lobes and eyes that are structurally similar to vertebrate species.
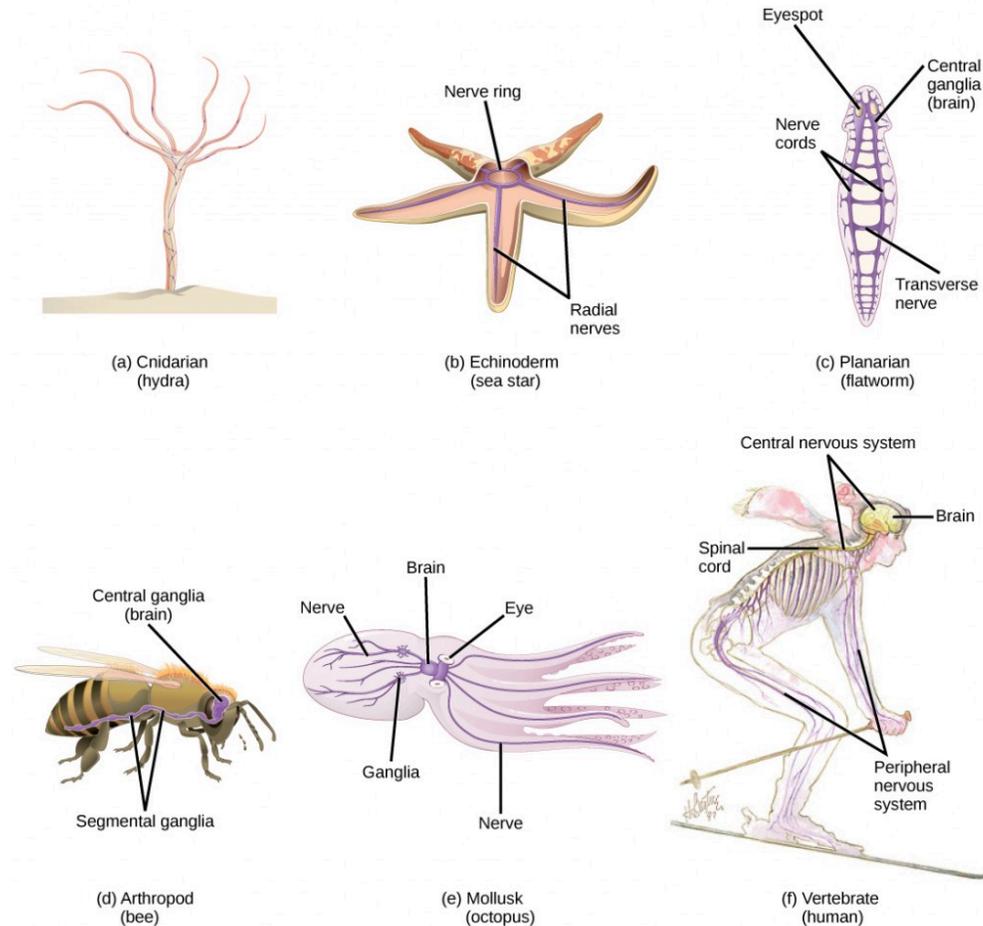
Figure 8.6. Nervous systems vary in structure and complexity. In (a) cnidarians, nerve cells form a decentralized nerve net. In (b) echinoderms, nerve cells are bundled into fibers called nerves. In animals exhibiting bilateral symmetry such as (c) planarians, neurons cluster into an anterior brain that processes information. In addition to a brain, (d) arthropods have clusters of nerve cell bodies, called peripheral ganglia, located along the ventral nerve cord. Mollusks such as squid and (e) octopi, which must hunt to survive, have complex brains containing millions of neurons. In (f) vertebrates, the brain and spinal cord comprise the central nervous system, while neurons extending into the rest of the body comprise the peripheral nervous system. (credit e: modification of work by Michael Vecchione, Clyde F.E. Roper, and Michael J. Sweeney, NOAA; credit f: modification of work by NIH)

Compared to invertebrates, vertebrate nervous systems are more complex, centralized, and specialized. While there is great diversity among different vertebrate nervous systems, they all share a basic structure: a CNS that contains a brain and spinal cord and a PNS made up of peripheral sensory and motor nerves. One interesting difference between the nervous systems of invertebrates and vertebrates is that the nerve cords of many invertebrates are located ventrally whereas the vertebrate spinal cords are located dorsally. There is debate among evolutionary biologists as to whether these different nervous system plans evolved separately or whether the invertebrate body plan arrangement somehow "flipped" during the evolution of vertebrates.

Watch [this video](#) of biologist Mark Kirschner discussing the "flipping" phenomenon of vertebrate evolution.

The nervous system is made up of **neurons**, specialized cells that can receive and transmit chemical or electrical signals, and **glia**, cells that provide support functions for the neurons by playing an information processing role that is complementary to neurons. A neuron can be compared to an electrical wire—it transmits a signal from one place to another. Glia can be compared to the workers at the electric company who make sure wires go to the right places, maintain the wires, and take down wires that are broken. Although glia have been compared to workers, recent evidence suggests that also usurp some of the signaling functions of neurons.

There is great diversity in the types of neurons and glia that are present in different parts of the nervous system. There are four major types of neurons, and they share several important cellular components.

Brain

The brain is the part of the central nervous system that is contained in the cranial cavity of the skull. It includes the cerebral cortex, limbic system, basal ganglia, thalamus, hypothalamus, and cerebellum. There are three different ways that a brain can be sectioned in order to view internal structures: a sagittal section cuts the brain left to right, as shown in [Figure 8.7](#), a coronal section cuts the brain front to back, as shown in [Figure 8.8](#), and a horizontal section cuts the brain top to bottom.
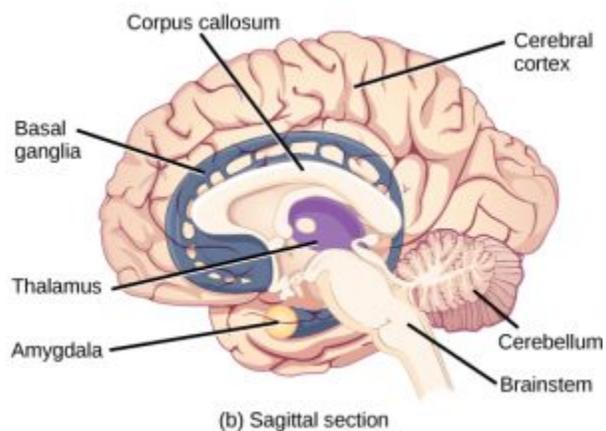

*Figure 8.7. Sagittal section of the human brain.*
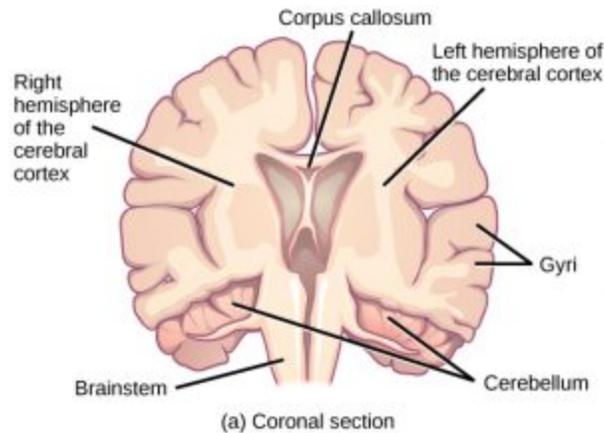
Figure 8.8. Coronal sections of the human brain.

Cerebral cortex

The outermost part of the brain is a thick piece of nervous system tissue called the **cerebral cortex**, which is folded into hills called **gyri** (singular: gyrus) and valleys called **sulci** (singular: sulcus). The cortex is made up of two hemispheres–right and left–which are separated by a large sulcus. A thick fiber bundle called the **corpus callosum** (Latin: "tough body") connects the two hemispheres and allows information to be passed from one side to the other. Although there are some brain functions that are localized more to one hemisphere than the other, the functions of the two hemispheres are largely redundant. In fact, sometimes (very rarely) an entire hemisphere is removed to treat severe epilepsy. While patients do suffer some deficits following the surgery, they can have surprisingly few problems, especially when the surgery is performed on children who have very immature nervous systems.

In other surgeries to treat severe epilepsy, the corpus callosum is cut instead of removing an entire hemisphere. This causes a condition called split-brain, which gives insights into unique functions of the two hemispheres. For example, when an object is presented to patients' left visual field, they may be unable to verbally name the object (and may claim to not have seen an object at all). This is because the visual input from the left visual field crosses and enters the right hemisphere and cannot then signal to the speech center, which generally is found in the left side of the brain. Remarkably, if a split-brain patient is asked to pick up a specific object out of a group of objects with the left hand, the patient will be able to do so but will still be unable to vocally identify it.

Each cortical hemisphere contains regions called lobes that are involved in different functions. Scientists use various techniques to determine what brain areas are involved in different functions: they examine patients who have had injuries or diseases that affect specific areas and see how those areas are related to functional deficits. They also conduct animal studies where they stimulate brain areas and see if there are any behavioral changes. They use a technique called transmagnetic stimulation (TMS) to temporarily deactivate specific parts of the cortex using strong magnets placed outside the head, and they use functional magnetic resonance imaging (fMRI) to look at changes in oxygenated blood flow in particular brain regions that correlate with specific behavioral tasks. These techniques have given great insight into the functions of different brain regions but have also shown that any given brain area can be involved in more than one behavior or process, and any given behavior or process generally involves neurons in multiple brain areas. That being said, each hemisphere of the mammalian cerebral cortex can be broken down into four functionally and spatially defined lobes: frontal, parietal, temporal, and occipital. [Figure 8.9](#) illustrates these four lobes of the human cerebral cortex.
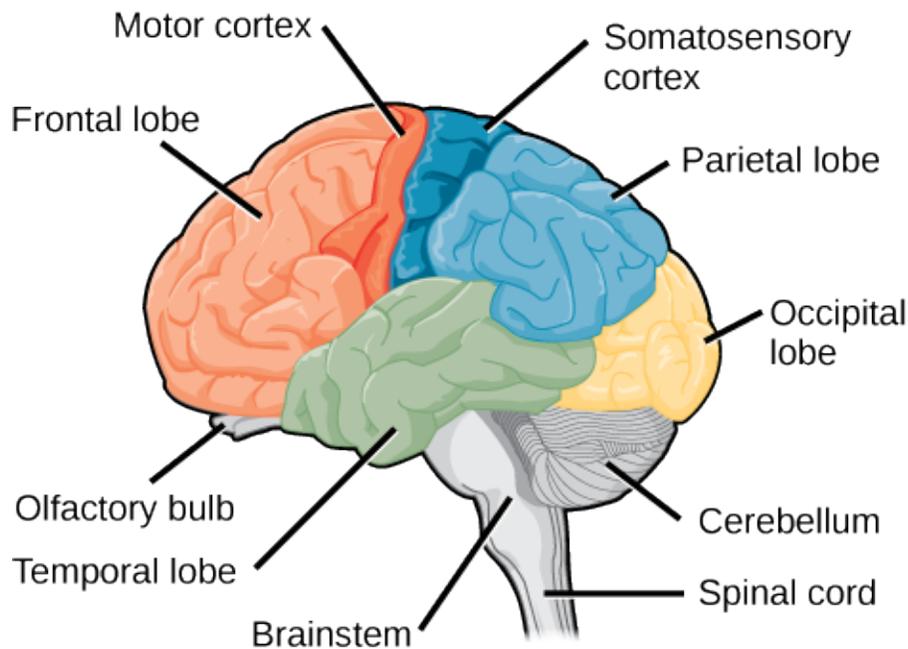

Figure 8.9. The human cerebral cortex includes the frontal, parietal, temporal, and occipital lobes.

The **frontal lobe** is located at the front of the brain, over the eyes. This lobe contains the olfactory bulb, which processes smells. The frontal lobe also contains the motor cortex, which is important for planning and implementing movement. Areas within the motor cortex map to different muscle groups and there is some organization to this map, as shown in Figure 8.10. For example, the neurons that control the movement of the fingers are next to the neurons that control the movement of the hand. Neurons in the frontal lobe also control cognitive functions like maintaining attention, speech, and decision-making. Studies of humans who have damaged their frontal lobes show that parts of this area are involved in personality, socialization, and assessing risk.
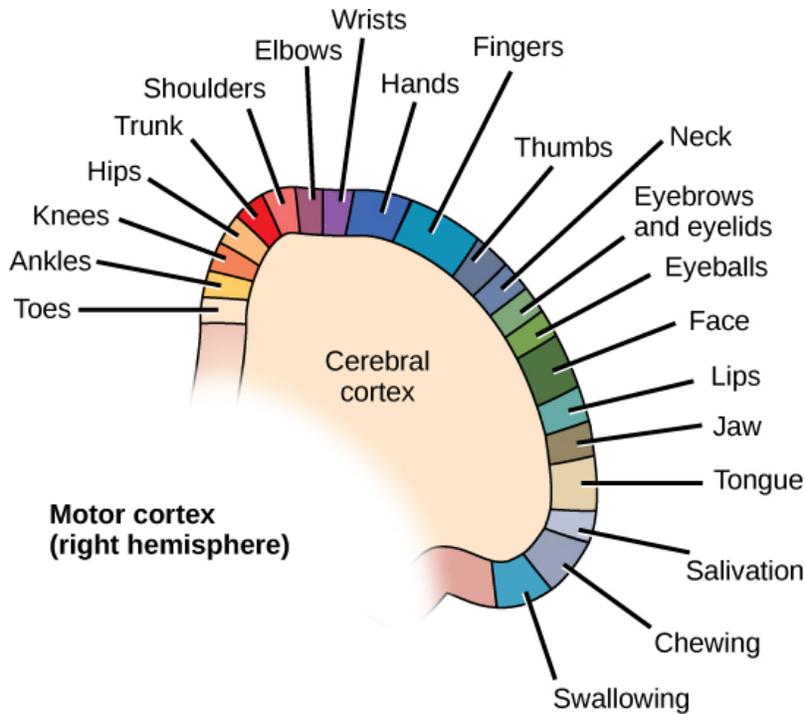
Figure 8.10. Different parts of the motor cortex control different muscle groups. Muscle groups that are neighbors in the body are generally controlled by neighboring regions of the motor cortex as well. For example, the neurons that control finger movement are near the neurons that control hand movement.

The **parietal lobe** is located at the top of the brain. Neurons in the parietal lobe are involved in speech and also reading. Two of the parietal lobe's main functions are processing **somatosensation**–touch sensations like pressure, pain, heat, cold–and processing **proprioception**–the sense of how parts of the body are oriented in space. The parietal lobe contains a somatosensory map of the body similar to the motor cortex.

The **occipital lobe** is located at the back of the brain. It is primarily involved in vision–seeing, recognizing, and identifying the visual world.

The **temporal lobe** is located at the base of the brain by your ears and is primarily involved in processing and interpreting sounds. It also contains the **hippocampus** (Greek for "seahorse")—a structure that processes memory formation. The hippocampus is illustrated in [Figure 8.9]. The role of the hippocampus in memory was partially determined by studying one famous epileptic patient, HM, who had both sides of his hippocampus removed in an attempt to cure his epilepsy. His seizures went away, but he could no longer form new memories (although he could remember some facts from before his surgery and could learn new motor tasks).

Compared to other vertebrates, mammals have exceptionally large brains for their body size. An entire alligator's brain, for example, would fill about one and a half teaspoons. This increase in brain to body size ratio is especially pronounced in apes, whales, and dolphins. While this increase in overall brain size doubtlessly played a role in the evolution of complex behaviors unique to mammals, it does not tell the whole story. Scientists have found a relationship between the relatively high surface area of the cortex and the intelligence and complex social behaviors exhibited by some mammals. This increased surface area is due, in part, to increased folding of the cortical sheet (more sulci and gyri). For example, a rat cortex is very smooth with very few sulci and gyri ([Figure 8.11]). Cat and sheep cortices have more sulci and gyri. Chimps, humans, and dolphins have even more.
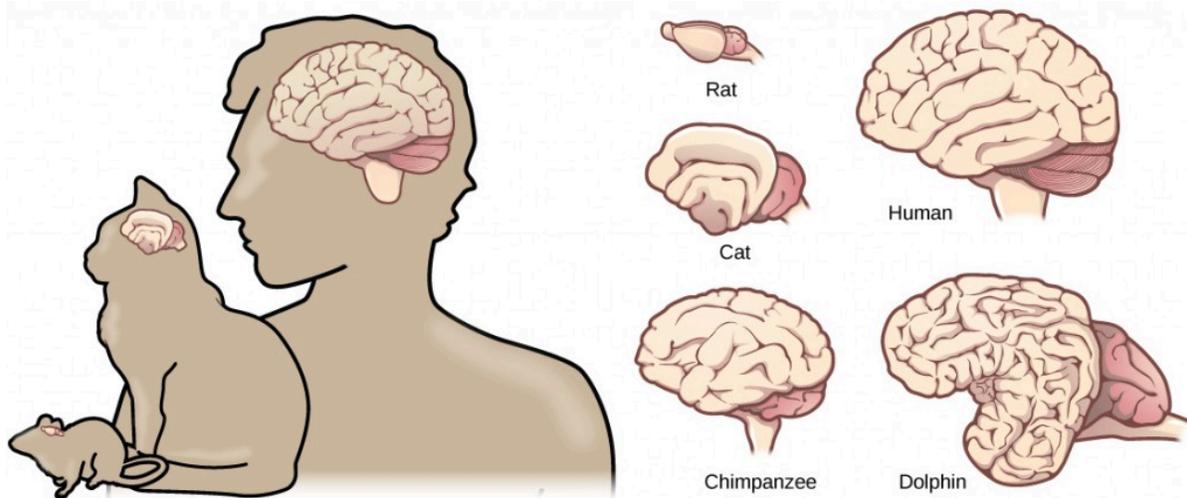

Figure 8.11. Mammals have larger brain-to-body ratios than other vertebrates. Within mammals, increased cortical folding and surface area is correlated with complex behavior.

Basal ganglia

Interconnected brain areas called the **basal ganglia (or basal nuclei),** shown in [Figure 8.7], play important roles in movement control and posture. Damage to the basal ganglia, as in Parkinson's disease, leads to motor impairments like a shuffling gait when walking. The basal ganglia also regulate motivation. For example, when a wasp sting led to bilateral basal ganglia damage in a 25-year-old

businessman, he began to spend all his days in bed and showed no interest in anything or anybody. But when he was externally stimulated–as when someone asked to play a card game with him–he was able to function normally. Interestingly, he and other similar patients do not report feeling bored or frustrated by their state.

Thalamus

The **thalamus** (Greek for "inner chamber"), illustrated in Figure 8.12, acts as a gateway to and from the cortex. It receives sensory and motor inputs from the body and also receives feedback from the cortex. This feedback mechanism can modulate conscious awareness of sensory and motor inputs depending on the attention and arousal state of the animal. The thalamus helps regulate consciousness, arousal, and sleep states. A rare genetic disorder called fatal familial insomnia causes the degeneration of thalamic neurons and glia. This disorder prevents affected patients from being able to sleep, among other symptoms, and is eventually fatal.
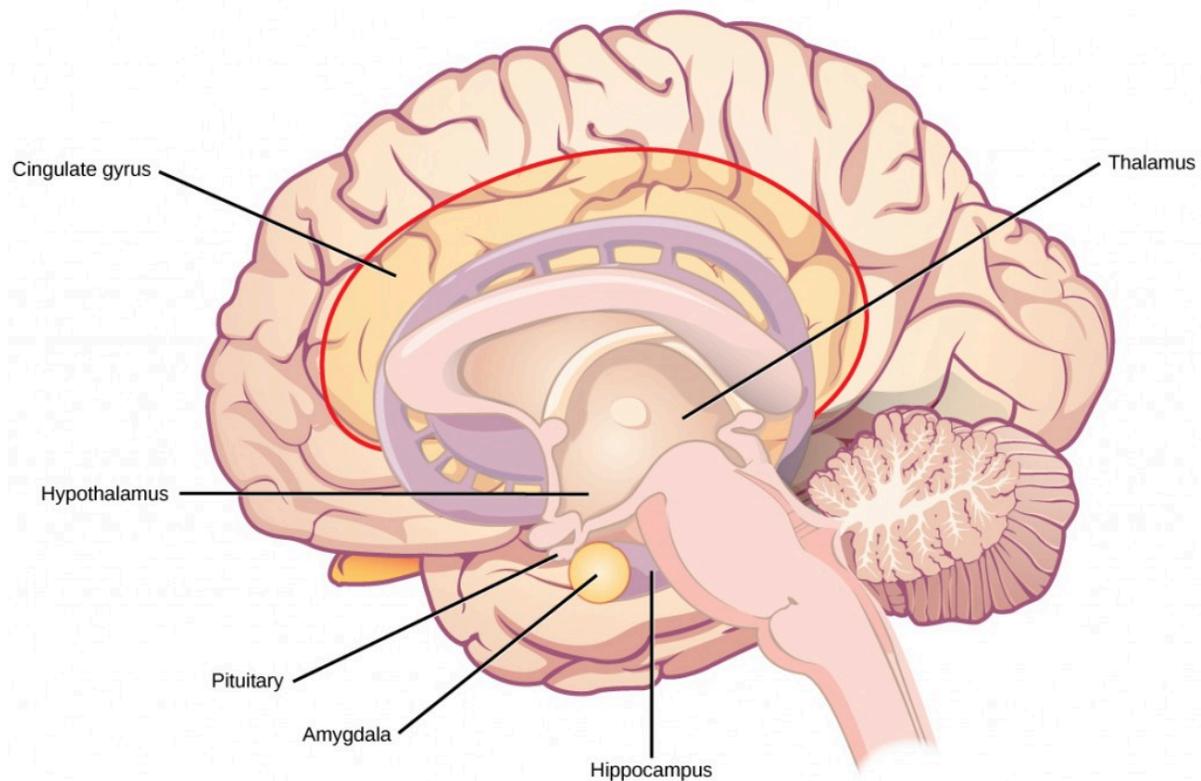

Figure 8.12. The limbic system regulates emotion and other behaviors. It includes parts of the cerebral cortex located near the center of the brain, including the cingulate gyrus and the hippocampus as well as the thalamus, hypothalamus, and amygdala.

Hypothalamus

Below the thalamus is the **hypothalamus**, shown in [Figure 8.12](#). The hypothalamus controls the endocrine system by sending signals to the pituitary gland, a pea-sized endocrine gland that releases several different hormones that affect other glands as well as other cells. This relationship means that the hypothalamus regulates important behaviors that are controlled by these hormones. The hypothalamus is the body's thermostat—it makes sure key functions like food and water intake, energy expenditure, and body temperature are kept at appropriate levels. Neurons within the hypothalamus also regulate circadian rhythms, sometimes called sleep cycles.

Limbic system

The **limbic system** is a connected set of structures that regulates emotion, as well as behaviors related to fear and motivation. It plays a role in memory formation and includes parts of the thalamus and hypothalamus as well as the hippocampus. One important structure within the limbic system is a temporal lobe structure called the **amygdala** (Greek for "almond"), illustrated in [Figure 8.12](#). The two amygdala are important both for the sensation of fear and for recognizing fearful faces. The **cingulate gyrus** helps regulate emotions and pain.

Cerebellum

The **cerebellum** (Latin for "little brain"), shown in [Figure 8.9](#), sits at the base of the brain on top of the brainstem. The cerebellum controls balance and aids in coordinating movement and learning new motor tasks.

Brainstem

The **brainstem**, illustrated in [Figure 8.9](#), connects the rest of the brain with the spinal cord. It consists of the midbrain, medulla oblongata, and the pons. Motor and sensory neurons extend through the brainstem allowing for the relay of signals between the brain and spinal cord. Ascending neural pathways cross in this section of the brain allowing the left hemisphere of the cerebrum to control the right side of the body and vice versa. The brainstem coordinates motor control signals sent from the brain to the body. The brainstem controls several important functions of the body including alertness, arousal, breathing, blood pressure, digestion, heart rate, swallowing, walking, and sensory and motor information integration.

Spinal cord

Connecting to the brainstem and extending down the body through the spinal column is the *spinal cord*, shown in [Figure 8.9](#). The spinal cord is a thick bundle of nerve tissue that carries information about the body to the brain and from the brain to the body. The spinal cord is contained within the bones of the vertebrate column but is able to communicate signals to and from the body through its connections with spinal nerves (part of the peripheral nervous system). A cross-section of the spinal cord looks like a white oval containing a gray butterfly-shape, as illustrated in [Figure 8.13](#). Myelinated axons make up the "white matter" and neuron and glial cell bodies make up the "gray matter." Gray matter is also composed of interneurons, which connect two neurons each located in different parts of the body. Axons and cell bodies in the dorsal (facing the back of the animal) spinal cord convey mostly sensory information from the body to the brain. Axons and cell bodies in the ventral (facing the front of the animal) spinal cord primarily transmit signals controlling movement from the brain to the body.

The spinal cord also controls motor reflexes. These reflexes are quick, unconscious movements—like automatically removing a hand from a hot object. Reflexes are so fast because they involve local synaptic connections. For example, the knee reflex that a doctor tests during a routine physical is controlled by a single synapse between a sensory neuron and a motor neuron. While a reflex may only require the involvement of one or two synapses, synapses with interneurons in the spinal column transmit information to the brain to convey what happened (the knee jerked, or the hand was hot).

In the United States, there around 10,000 spinal cord injuries each year. Because the spinal cord is the information superhighway connecting the brain with the body, damage to the spinal cord can lead to paralysis. The extent of the paralysis depends on the location of the injury along the spinal cord and whether the spinal cord was completely severed. For example, if the spinal cord is damaged at the level of the neck, it can cause paralysis from the neck down, whereas damage to the spinal column further down may limit paralysis to the legs. Spinal cord injuries are notoriously difficult to treat because spinal nerves do not regenerate, although ongoing research suggests that stem cell transplants may be able to act as a bridge to reconnect severed nerves. Researchers are also looking at ways to prevent the inflammation that worsens nerve damage after injury. One such treatment is to pump the body with cold saline to induce hypothermia. This cooling can prevent swelling and other processes that are thought to worsen spinal cord injuries.
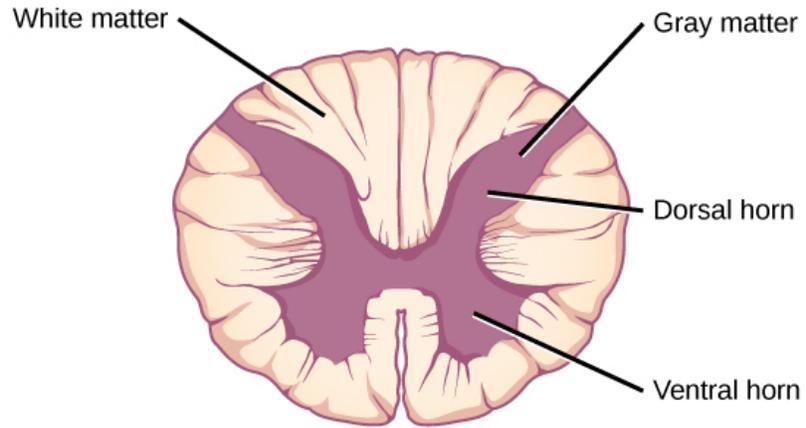
Figure 8.13. A cross-section of the spinal cord shows gray matter (containing cell bodies and interneurons) and white matter (containing axons).

Question 8.3
The _____ lobe contains the visual cortex.
a. frontal
b. parietal
c. temporal
d. occipital

Question 8.4
The _____ connects the two cerebral hemispheres.
a. limbic system
b. corpus callosum
c. cerebellum
d. pituitary

Question 8.5
Neurons in the _____ control motor reflexes.
a. thalamus
b. spinal cord
c. parietal lobe
d. hippocampus

Question 8.6
What methods can be used to determine the function of a particular brain region?

Question 8.7
What are the main functions of the spinal cord?

Question 8.8
Which brain structure is the oldest on the evolutionary tree of animals? Which animal groups have this structure?

8.3 – Special Senses

8.3. Explain how olfactory, gustatory, equilibrium, auditory, and visual sensations are interpreted by animals with reference to signal processing and/or specific ion movements.

Sensory information processing in animals

In more advanced animals, the senses are constantly at work, making the animal aware of stimuli—such as light, or sound, or the presence of a chemical substance in the external environment—and monitoring information about the organism's internal environment. All bilaterally symmetric animals have a sensory system, and the development of any species' sensory system has been driven by natural selection; thus, sensory systems differ among species according to the demands of their environments. The shark, unlike most fish predators, is electrosensitive—that is, sensitive to electrical fields produced by other animals in its environment. While it is helpful to this underwater predator, electrosensitivity is a sense not found in most land animals.

Senses provide information about the body and its environment. Humans have five special senses: olfaction (smell), gustation (taste), equilibrium (balance and body position), vision, and hearing. Additionally, we possess general senses, also called somatosensation, which respond to stimuli like temperature, pain, pressure, and vibration. **Vestibular sensation**, which is an organism's sense of spatial orientation and balance, **proprioception** (position of bones, joints, and muscles), and the sense of limb position that is used to track **kinesthesia** (limb movement) are part of somatosensation. Although the sensory systems associated with these senses are very different, all share a common function: to convert a stimulus (such as light, or sound, or the position of the body) into an electrical signal in the nervous system. This process is called **sensory transduction**.

There are two broad types of cellular systems that perform sensory transduction. In one, a neuron works with a **sensory receptor**, a cell, or cell process that is specialized to engage with and detect a specific stimulus. Stimulation of the sensory receptor activates the associated afferent neuron, which carries information about the stimulus to the central nervous system. In the second type of sensory transduction, a sensory nerve ending responds to a stimulus in the internal or external environment: this neuron constitutes the sensory receptor. Free nerve endings can be stimulated by

several different stimuli, thus showing little receptor specificity. For example, pain receptors in your gums and teeth may be stimulated by temperature changes, chemical stimulation, or pressure.

Reception

The first step in sensation is **reception**, which is the activation of sensory receptors by stimuli such as mechanical stimuli (being bent or squished, for example), chemicals, or temperature. The receptor can then respond to the stimuli. The region in space in which a given sensory receptor can respond to a stimulus, be it far away or in contact with the body, is that receptor's **receptive** field. Think for a moment about the differences in receptive fields for the different senses. For the sense of touch, a stimulus must come into contact with body. For the sense of hearing, a stimulus can be a moderate distance away (some baleen whale sounds can propagate for many kilometers). For vision, a stimulus can be very far away; for example, the visual system perceives light from stars at enormous distances.

Transduction

The most fundamental function of a sensory system is the translation of a sensory signal to an electrical signal in the nervous system. This takes place at the sensory receptor, and the change in electrical potential that is produced is called the **receptor potential**. How is sensory input, such as pressure on the skin, changed to a receptor potential? In this example, a type of receptor called a **mechanoreceptor** (as shown in Figure 8.14) possesses specialized membranes that respond to pressure. Disturbance of these dendrites by compressing them or bending them opens gated ion channels in the plasma membrane of the sensory neuron, changing its electrical potential. Recall that in the nervous system, a positive change of a neuron's electrical potential (also called the membrane potential), depolarizes the neuron. Receptor potentials are graded potentials: the magnitude of these graded (receptor) potentials varies with the strength of the stimulus. If the magnitude of depolarization is sufficient (that is, if membrane potential reaches a threshold), the neuron will fire an action potential. In most cases, the correct stimulus impinging on a sensory receptor will drive membrane potential in a positive direction, although for some receptors, such as those in the visual system, this is not always the case.
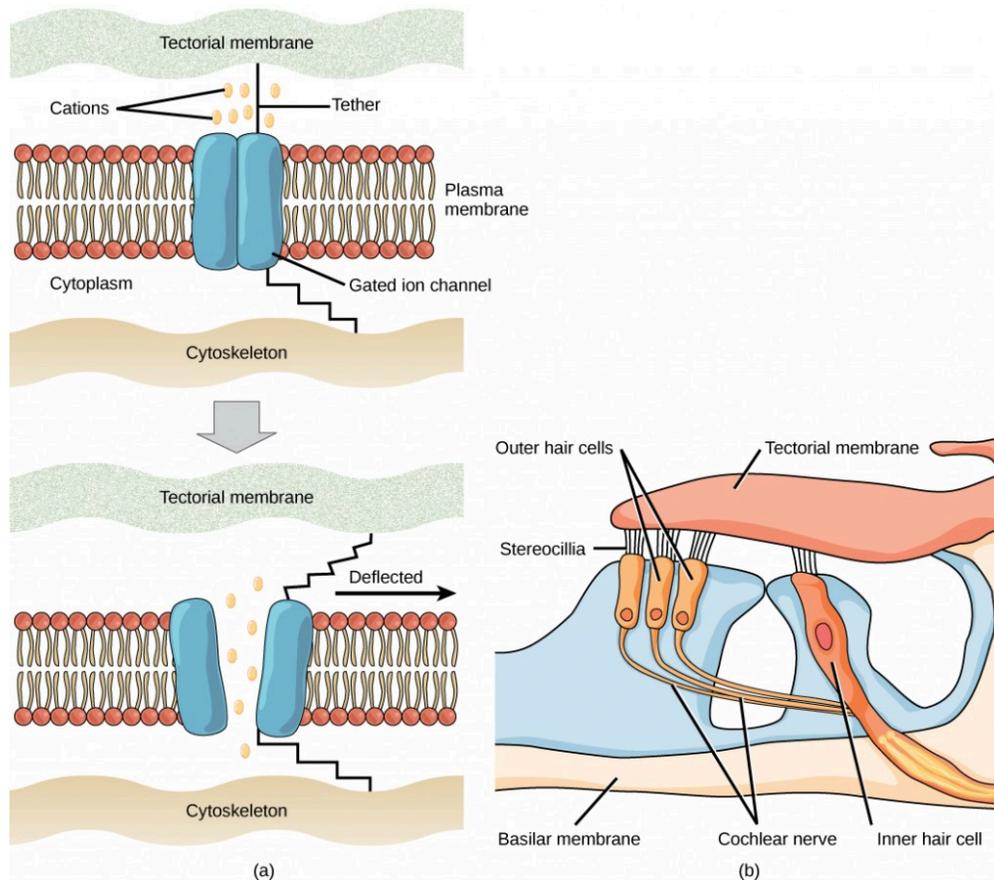
Figure 8.14. (a) Mechanosensitive ion channels are gated ion channels that respond to mechanical deformation of the plasma membrane. A mechanosensitive channel is connected to the plasma membrane and the cytoskeleton by hair-like tethers. When pressure causes the extracellular matrix to move, the channel opens, allowing ions to enter or exit the cell. (b) Stereocilia in the human ear are connected to mechanosensitive ion channels. When a sound causes the stereocilia to move, mechanosensitive ion channels transduce the signal to the cochlear nerve.

Sensory receptors for different senses are very different from each other, and they are specialized according to the type of stimulus they sense: they have receptor specificity. For example, touch receptors, light receptors, and sound receptors are each activated by different stimuli. Touch receptors are not sensitive to light or sound; they are sensitive only to touch or pressure. However, stimuli may be combined at higher levels in the brain, as happens with olfaction, contributing to our sense of taste.

Encoding and transmission of sensory information

Four aspects of sensory information are encoded by sensory systems: the type of stimulus, the

location of the stimulus in the receptive field, the duration of the stimulus, and the relative intensity of the stimulus. Thus, action potentials transmitted over a sensory receptor's afferent axons encode one type of stimulus, and this segregation of the senses is preserved in other sensory circuits. For example, auditory receptors transmit signals over their own dedicated system, and electrical activity in the axons of the auditory receptors will be interpreted by the brain as an auditory stimulus–a sound.

The intensity of a stimulus is often encoded in the rate of action potentials produced by the sensory receptor. Thus, an intense stimulus will produce a more rapid train of action potentials, and reducing the stimulus will likewise slow the rate of production of action potentials. A second way in which intensity is encoded is by the number of receptors activated. An intense stimulus might initiate action potentials in a large number of adjacent receptors, while a less intense stimulus might stimulate fewer receptors. Integration of sensory information begins as soon as the information is received in the CNS, and the brain will further process incoming signals.

Perception

Perception is an individual's interpretation of a sensation. Although perception relies on the activation of sensory receptors, perception happens not at the level of the sensory receptor, but at higher levels in the nervous system, in the brain. The brain distinguishes sensory stimuli through a sensory pathway: action potentials from sensory receptors travel along neurons that are dedicated to a particular stimulus. These neurons are dedicated to that particular stimulus and synapse with particular neurons in the brain or spinal cord.

All sensory signals, except those from the olfactory system, are transmitted though the central nervous system and are routed to the thalamus and to the appropriate region of the cortex. Recall that the thalamus is a structure in the forebrain that serves as a clearinghouse and relay station for sensory (as well as motor) signals. When the sensory signal exits the thalamus, it is conducted to the specific area of the cortex (Figure 8.15) dedicated to processing that particular sense.

How are neural signals interpreted? Interpretation of sensory signals between individuals of the same species is largely similar, owing to the inherited similarity of their nervous systems; however, there are some individual differences. A good example of this is individual tolerances to a painful stimulus, such as dental pain, which certainly differ.
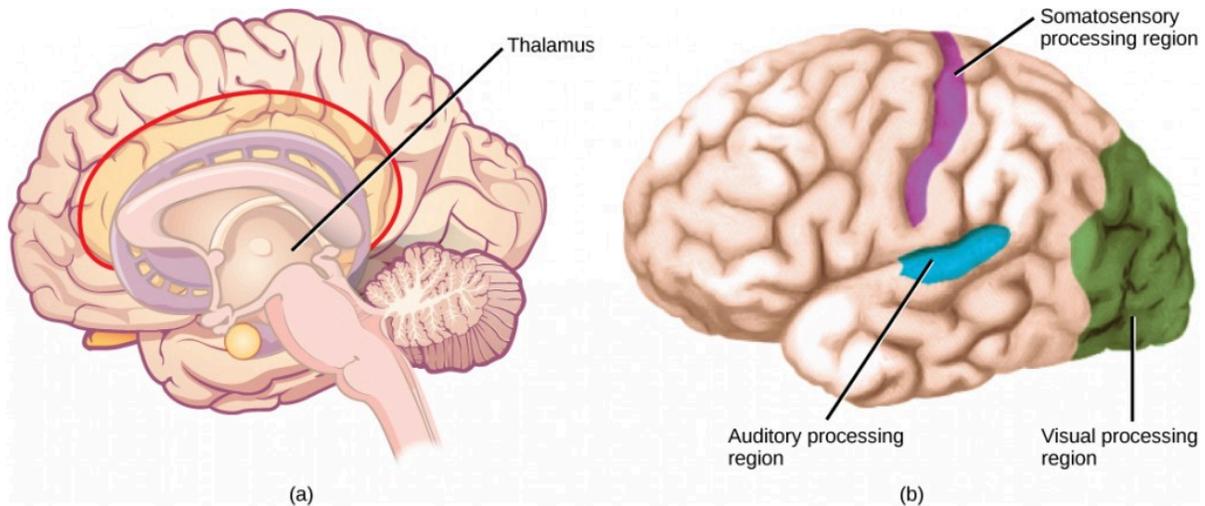
Figure 8.15. *In humans, with the exception of olfaction, all sensory signals are routed from the (a) thalamus to (b) final processing regions in the cortex of the brain. (credit b: modification of work by Polina Tishina) Scientific Method Connection*

Question 8.9

Where does perception occur?
a. spinal cord
b. cerebral cortex
c. receptors
d. thalamus

Question 8.10

If a person's cold receptors no longer convert cold stimuli into sensory signals, that person has a problem with the process of _____.
a. reception
b. transmission
c. perception
d. transduction

Question 8.11

After somatosensory transduction, the sensory signal travels through the brain as a(n) _____ signal.
a. electrical
b. pressure
c. optical
d. thermal

Somatosensation

Somatosensation is a mixed sensory category and includes all sensation received from the skin and mucous membranes, as well from as the limbs and joints. Somatosensation is also known as tactile sense, or more familiarly, as the sense of touch. Somatosensation occurs all over the exterior of the body and at some interior locations as well. A variety of receptor types—embedded in the skin, mucous membranes, muscles, joints, internal organs, and cardiovascular system—play a role.

Somatosensory receptors

Sensory receptors are classified into five categories: mechanoreceptors, thermoreceptors, proprioceptors, pain receptors, and chemoreceptors. These categories are based on the nature of stimuli each receptor class transduces. What is commonly referred to as "touch" involves more than one kind of stimulus and more than one kind of receptor. Mechanoreceptors in the skin are described as encapsulated (that is, surrounded by a capsule) or unencapsulated (a group that includes free nerve endings). A **free nerve ending**, as its name implies, is an unencapsulated dendrite of a sensory neuron. Free nerve endings are the most common nerve endings in skin, and they extend into the middle of the epidermis. Free nerve endings are sensitive to painful stimuli, to hot and cold, and to light touch. They are slow to adjust to a stimulus and so are less sensitive to abrupt changes in stimulation.

There are three classes of mechanoreceptors: tactile, proprioceptors, and baroreceptors. Mechanoreceptors sense stimuli due to physical deformation of their plasma membranes. They contain mechanically gated ion channels whose gates open or close in response to pressure, touch, stretching, and sound." There are four primary tactile mechanoreceptors in human skin: Merkel's disks, Meissner's corpuscles, Ruffini endings, and Pacinian corpuscle; two are located toward the surface of the skin and two are located deeper. A fifth type of mechanoreceptor, Krause end bulbs, are found only in specialized regions. **Merkel's disks** (shown in [Figure 8.16](#)) are found in the upper layers of skin near the base of the epidermis, both in skin that has hair and on **glabrous** skin, that is, the hairless skin found on the palms and fingers, the soles of the feet, and the lips of humans and other primates. Merkel's disks are densely distributed in the fingertips and lips. They are slow-adapting, unencapsulated nerve endings, and they respond to light touch. Light touch, also known as discriminative touch, is a light pressure that allows the location of a stimulus to be pinpointed. The receptive fields of Merkel's disks are small with well-defined borders. That makes them finely sensitive to edges and they come into use in tasks such as typing on a keyboard.
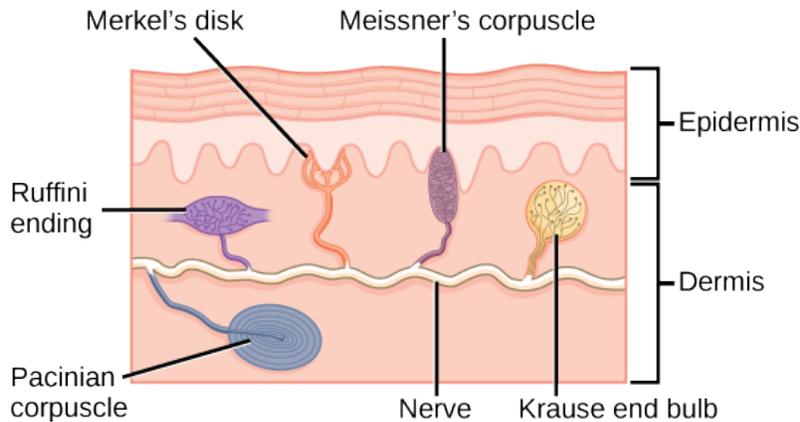
Figure 8.16. Four of the primary mechanoreceptors in human skin are shown. Merkel's disks, which are unencapsulated, respond to light touch. Meissner's corpuscles, Ruffini endings, Pacinian corpuscles, and Krause end bulbs are all encapsulated. Meissner's corpuscles respond to touch and low-frequency vibration. Ruffini endings detect stretch, deformation within joints, and warmth. Pacinian corpuscles detect transient pressure and high-frequency vibration. Krause end bulbs detect cold.

Question 8.12

Which of the following statements about mechanoreceptors is false?
a. Pacini corpuscles are found in both glabrous and hairy skin.
b. Merkel's disks are abundant on the fingertips and lips.
c. Ruffini endings are encapsulated mechanoreceptors.
d. Meissner's corpuscles extend into the lower dermis.

In proprioception, proprioceptive and kinesthetic signals travel through myelinated afferent neurons running from the spinal cord to the medulla. Neurons are not physically connected, but communicate via neurotransmitters secreted into synapses or "gaps" between communicating neurons. Once in the medulla, the neurons continue carrying the signals to the thalamus.

Muscle spindles are stretch receptors that detect the amount of stretch, or lengthening of muscles. Related to these are **Golgi tendon organs**, which are tension receptors that detect the force of muscle contraction. Proprioceptive and kinesthetic signals come from limbs. Unconscious proprioceptive signals run from the spinal cord to the cerebellum, the brain region that coordinates muscle contraction, rather than to the thalamus, like most other sensory information.

Barorecptors detect pressure changes in an organ. They are found in the walls of the carotid artery and the aorta where they monitor blood pressure, and in the lungs where they detect the degree of lung expansion. Stretch receptors are found at various sites in the digestive and urinary systems.

In addition to these two types of deeper receptors, there are also rapidly adapting hair receptors, which are found on nerve endings that wrap around the base of hair follicles. There are a few types of hair receptors that detect slow and rapid hair movement, and they differ in their sensitivity to movement. Some hair receptors also detect skin deflection, and certain rapidly adapting hair receptors allow detection of stimuli that have not yet touched the skin.

Integration of signals from mechanoreceptors

The configuration of the different types of receptors working in concert in human skin results in a very refined sense of touch. The nociceptive receptors–those that detect pain–are located near the surface. Small, finely calibrated mechanoreceptors–Merkel's disks and Meissner's corpuscles–are located in the upper layers and can precisely localize even gentle touch. The large mechanoreceptors–Pacinian corpuscles and Ruffini endings–are located in the lower layers and respond to deeper touch. (Consider that the deep pressure that reaches those deeper receptors would not need to be finely localized.) Both the upper and lower layers of the skin hold rapidly and slowly adapting receptors. Both primary somatosensory cortex and secondary cortical areas are responsible for processing the complex picture of stimuli transmitted from the interplay of mechanoreceptors.

Density of mechanoreceptors

The distribution of touch receptors in human skin is not consistent over the body. In humans, touch receptors are less dense in skin covered with any type of hair, such as the arms, legs, torso, and face. Touch receptors are denser in glabrous skin (the type found on human fingertips and lips, for example), which is typically more sensitive and is thicker than hairy skin (4 to 5 mm versus 2 to 3 mm).

How is receptor density estimated in a human subject? The relative density of pressure receptors in different locations on the body can be demonstrated experimentally using a two-point discrimination test. In this demonstration, two sharp points, such as two thumbtacks, are brought into contact with the subject's skin (though not hard enough to cause pain or break the skin). The subject reports if he or she feels one point or two points. If the two points are felt as one point, it can be inferred that the two points are both in the receptive field of a single sensory receptor. If two points are felt as two separate points, each is in the receptive field of two separate sensory receptors. The points could

then be moved closer and re-tested until the subject reports feeling only one point, and the size of the receptive field of a single receptor could be estimated from that distance.

Thermoreception

In addition to Krause end bulbs that detect cold and Ruffini endings that detect warmth, there are different types of cold receptors on some free nerve endings: thermoreceptors, located in the dermis, skeletal muscles, liver, and hypothalamus, that are activated by different temperatures. Their pathways into the brain run from the spinal cord through the thalamus to the primary somatosensory cortex. Warmth and cold information from the face travels through one of the cranial nerves to the brain. You know from experience that a tolerably cold or hot stimulus can quickly progress to a much more intense stimulus that is no longer tolerable. Any stimulus that is too intense can be perceived as pain because temperature sensations are conducted along the same pathways that carry pain sensations

Pain

Pain is the name given to **nociception**, which is the neural processing of injurious stimuli in response to tissue damage. Pain is caused by true sources of injury, such as contact with a heat source that causes a thermal burn or contact with a corrosive chemical. But pain also can be caused by harmless stimuli that mimic the action of damaging stimuli, such as contact with capsaicins, the compounds that cause peppers to taste hot and which are used in self-defense pepper sprays and certain topical medications. Peppers taste "hot" because the protein receptors that bind capsaicin open the same calcium channels that are activated by warm receptors.

Nociception starts at the sensory receptors, but pain, inasmuch as it is the perception of nociception, does not start until it is communicated to the brain. There are several nociceptive pathways to and through the brain. Most axons carrying nociceptive information into the brain from the spinal cord project to the thalamus (as do other sensory neurons) and the neural signal undergoes final processing in the primary somatosensory cortex. Interestingly, one nociceptive pathway projects not to the thalamus but directly to the hypothalamus in the forebrain, which modulates the cardiovascular and neuroendocrine functions of the autonomic nervous system. Recall that threatening–or painful–stimuli stimulate the sympathetic branch of the visceral sensory system, readying a fight-or-flight response.

View this [video](#) that animates the five phases of nociceptive pain.

Taste and smell

Taste, also called **gustation**, and smell, also called **olfaction**, are the most interconnected senses in that both involve molecules of the stimulus entering the body and bonding to receptors. Smell lets an animal sense the presence of food or other animals—whether potential mates, predators, or prey—or other chemicals in the environment that can impact their survival. Similarly, the sense of taste allows animals to discriminate between types of foods. While the value of a sense of smell is obvious, what is the value of a sense of taste? Different tasting foods have different attributes, both helpful and harmful. For example, sweet-tasting substances tend to be highly caloric, which could be necessary for survival in lean times. Bitterness is associated with toxicity, and sourness is associated with spoiled food. Salty foods are valuable in maintaining homeostasis by helping the body retain water and by providing ions necessary for cells to function.

Tastes and odors

Both taste and odor stimuli are molecules taken in from the environment. The primary tastes detected by humans are sweet, sour, bitter, salty and umami. The first four tastes need little explanation. The identification of **umami** as a fundamental taste occurred fairly recently—it was identified in 1908 by Japanese scientist Kikunae Ikeda while he worked with seaweed broth, but it was not widely accepted as a taste that could be physiologically distinguished until many years later. The taste of umami, also known as savoriness, is attributable to the taste of the amino acid L-glutamate. In fact, monosodium glutamate, or MSG, is often used in cooking to enhance the savory taste of certain foods. What is the adaptive value of being able to distinguish umami? Savory substances tend to be high in protein.

All odors that we perceive are molecules in the air we breathe. If a substance does not release molecules into the air from its surface, it has no smell. And if a human or other animal does not have a receptor that recognizes a specific molecule, then that molecule has no smell. Humans have about 350 olfactory receptor subtypes that work in various combinations to allow us to sense about 10,000

different odors. Compare that to mice, for example, which have about 1,300 olfactory receptor types, and therefore probably sense more odors. Both odors and tastes involve molecules that stimulate specific chemoreceptors. Although humans commonly distinguish taste as one sense and smell as another, they work together to create the perception of flavor. A person's perception of flavor is reduced if he or she has congested nasal passages.

Reception and transduction

Odorants (odor molecules) enter the nose and dissolve in the olfactory epithelium, the mucosa at the back of the nasal cavity (as illustrated in [Figure 8.17](#)). The **olfactory epithelium** is a collection of specialized olfactory receptors in the back of the nasal cavity that spans an area about 5 cm^2 in humans. Recall that sensory cells are neurons. An **olfactory receptor**, which is a dendrite of a specialized neuron, responds when it binds certain molecules inhaled from the environment by sending impulses directly to the olfactory bulb of the brain. Humans have about 12 million olfactory receptors, distributed among hundreds of different receptor types that respond to different odors. Twelve million seems like a large number of receptors, but compare that to other animals: rabbits have about 100 million, most dogs have about 1 billion, and bloodhounds–dogs selectively bred for their sense of smell–have about 4 billion. The overall size of the olfactory epithelium also differs between species, with that of bloodhounds, for example, being many times larger than that of humans.

Olfactory neurons are **bipolar neurons** (neurons with two processes from the cell body). Each neuron has a single dendrite buried in the olfactory epithelium, and extending from this dendrite are 5 to 20 receptor-laden, hair-like cilia that trap odorant molecules. The sensory receptors on the cilia are proteins, and it is the variations in their amino acid chains that make the receptors sensitive to different odorants. Each olfactory sensory neuron has only one type of receptor on its cilia, and the receptors are specialized to detect specific odorants, so the bipolar neurons themselves are specialized. When an odorant binds with a receptor that recognizes it, the sensory neuron associated with the receptor is stimulated. Olfactory stimulation is the only sensory information that directly reaches the cerebral cortex, whereas other sensations are relayed through the thalamus.
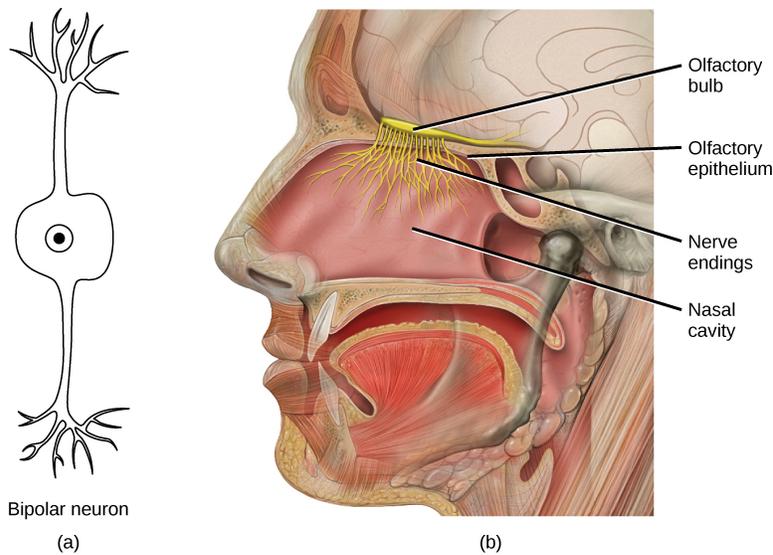
Figure 8.17. In the human olfactory system, (a) bipolar olfactory neurons extend from (b) the olfactory epithelium, where olfactory receptors are located, to the olfactory bulb. (credit: modification of work by Patrick J. Lynch, medical illustrator; C. Carl Jaffe, MD, cardiologist)

Pheromones

A **pheromone** is a chemical released by an animal that affects the behavior or physiology of animals of the same species. Pheromonal signals can have profound effects on animals that inhale them, but pheromones apparently are not consciously perceived in the same way as other odors. There are several different types of pheromones, which are released in urine or as glandular secretions. Certain pheromones are attractants to potential mates, others are repellants to potential competitors of the same sex, and still others play roles in mother-infant attachment. Some pheromones can also influence the timing of puberty, modify reproductive cycles, and even prevent embryonic implantation. While the roles of pheromones in many nonhuman species are important, pheromones have become less important in human behavior over evolutionary time compared to their importance to organisms with more limited behavioral repertoires.

The vomeronasal organ (VNO, or Jacobson's organ) is a tubular, fluid-filled, olfactory organ present in many vertebrate animals that sits adjacent to the nasal cavity. It is very sensitive to pheromones and is connected to the nasal cavity by a duct. When molecules dissolve in the mucosa of the nasal cavity, they then enter the VNO where the pheromone molecules among them bind with specialized pheromone receptors. Upon exposure to pheromones from their own species or others, many animals, including cats, may display the flehmen response (shown in [Figure 8.18](#)), a curling of the upper lip that helps pheromone molecules enter the VNO.

Pheromonal signals are sent, not to the main olfactory bulb, but to a different neural structure that

projects directly to the amygdala (recall that the amygdala is a brain center important in emotional reactions, such as fear). The pheromonal signal then continues to areas of the hypothalamus that are key to reproductive physiology and behavior. While some scientists assert that the VNO is apparently functionally vestigial in humans, eve though there is a similar structure located near human nasal cavities, others are researching it as a possible functional system that may, for example, contribute to synchronization of menstrual cycles in women living in close proximity.

Figure 8.18. The flehmen response in this tiger results in the curling of the upper lip and helps airborne pheromone molecules enter the vomeronasal organ. (credit: modification of work by "chadh"/Flickr)

Taste

Detecting a taste (gustation) is fairly similar to detecting an odor (olfaction), given that both taste and smell rely on chemical receptors being stimulated by certain molecules. The primary organ of taste is the taste bud. A *taste bud* is a cluster of gustatory receptors (taste cells) that are located within the bumps on the tongue called *papillae* (singular: papilla) (illustrated in [Figure 8.19](#)). There are several structurally distinct papillae. Filiform papillae, which are located across the tongue, are tactile, providing friction that helps the tongue move substances, and contain no taste cells. In contrast, fungiform papillae, which are located mainly on the anterior two-thirds of the tongue, each contain one to eight taste buds and also have receptors for pressure and temperature. The large circumvallate papillae contain up to 100 taste buds and form a V near the posterior margin of the tongue.
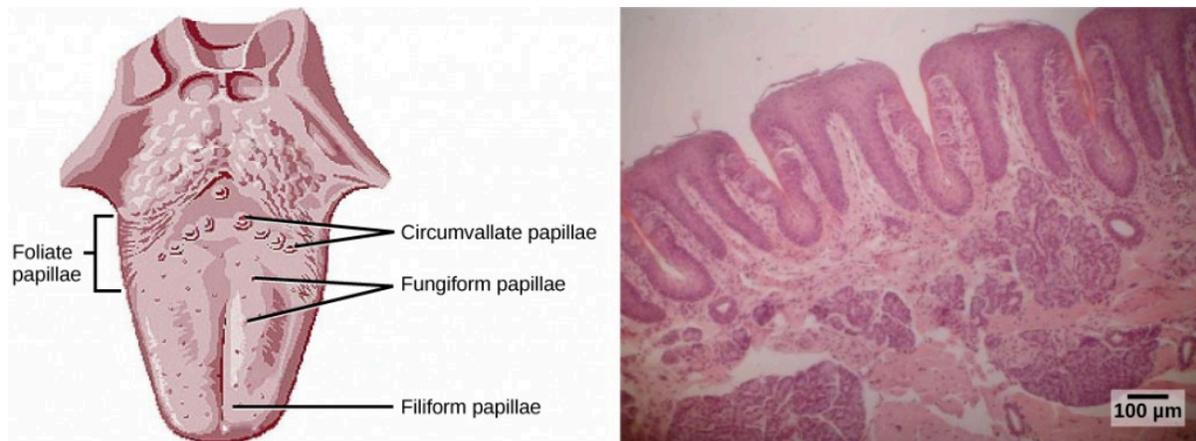
Figure 8.19. (a) Foliate, circumvallate, and fungiform papillae are located on different regions of the tongue. (b) Foliate papillae are prominent protrusions on this light micrograph. (credit a: modification of work by NCI; scale-bar data from Matt Russell)

In addition to those two types of chemically and mechanically sensitive papillae are foliate papillae—leaf-like papillae located in parallel folds along the edges and toward the back of the tongue, as seen in the Figure 8.18b micrograph. Foliate papillae contain about 1,300 taste buds within their folds. Finally, there are circumvallate papillae, which are wall-like papillae in the shape of an inverted "V" at the back of the tongue. Each of these papillae is surrounded by a groove and contains about 250 taste buds.

Each taste bud's taste cells are replaced every 10 to 14 days. These are elongated cells with hair-like processes called microvilli at the tips that extend into the taste bud pore (illustrate in Figure 8.20). Food molecules (**tastants**) are dissolved in saliva, and they bind with and stimulate the receptors on the microvilli. The receptors for tastants are located across the outer portion and front of the tongue, outside of the middle area where the filiform papillae are most prominent.
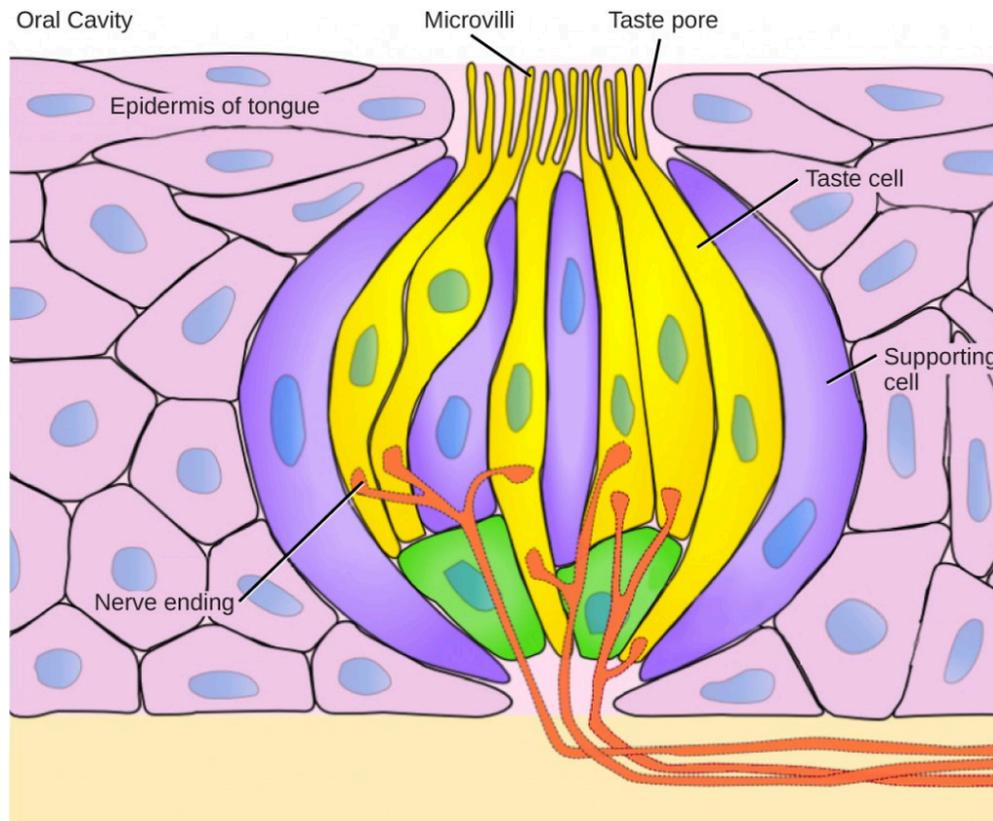
Figure 8.20. Pores in the tongue allow tastants to enter taste pores in the tongue. (credit: modification of work by Vincenzo Rizzo)

In humans, there are five primary tastes, and each taste has only one corresponding type of receptor. Thus, like olfaction, each receptor is specific to its stimulus (tastant). Transduction of the five tastes happens through different mechanisms that reflect the molecular composition of the tastant. A salty tastant (containing NaCl) provides the sodium ions (Na$^+$) that enter the taste neurons and excite them directly. Sour tastants are acids and belong to the thermoreceptor protein family. Binding of an acid or other sour-tasting molecule triggers a change in the ion channel and these increase hydrogen ion (H$^+$) concentrations in the taste neurons, thus depolarizing them. Sweet, bitter, and umami tastants require a G-protein coupled receptor. These tastants bind to their respective receptors, thereby exciting the specialized neurons associated with them. Both tasting abilities and sense of smell change with age. In humans, the senses decline dramatically by age 50 and continue to decline. A child may find a food to be too spicy, whereas an elderly person may find the same food to be bland and unappetizing.

View this [animation](#) that shows how the sense of taste works.

Smell and taste in the brain

Olfactory neurons project from the olfactory epithelium to the olfactory bulb as thin, unmyelinated axons. The **olfactory bulb** is composed of neural clusters called **glomeruli**, and each glomerulus receives signals from one type of olfactory receptor, so each glomerulus is specific to one odorant. From glomeruli, olfactory signals travel directly to the olfactory cortex and then to the frontal cortex and the thalamus. Recall that this is a different path from most other sensory information, which is sent directly to the thalamus before ending up in the cortex. Olfactory signals also travel directly to the amygdala, thereafter reaching the hypothalamus, thalamus, and frontal cortex. The last structure that olfactory signals directly travel to is a cortical center in the temporal lobe structure important in spatial, autobiographical, declarative, and episodic memories. Olfaction is finally processed by areas of the brain that deal with memory, emotions, reproduction, and thought.

Taste neurons project from taste cells in the tongue, esophagus, and palate to the medulla, in the brainstem. From the medulla, taste signals travel to the thalamus and then to the primary gustatory cortex. Information from different regions of the tongue is segregated in the medulla, thalamus, and cortex.

Question 8.13

Salty foods activate the taste cells by _ _ _ _ _ _.
a. exciting the taste cell directly
b. causing hydrogen ions to enter the cell
d. causing sodium channels to close
e. binding directly to the receptors

Question 8.14

All sensory signals except _ _ _ _ _ travel to the _ _ _ _ _ in the brain before the cerebral cortex.
a. vision; thalamus
b. olfaction; thalmus
c. vision; cranial nerves
d. olfaction; cranial nerves

Question 8.15

From the perspective of the recipient of the signal, in what ways do pheromones differ from other odorants?

Question 8.16

What might be the effect on an animal of not being able to perceive taste?

Hearing and vestibular information

Audition, or hearing, is important to humans and to other animals for many different interactions. It enables an organism to detect and receive information about danger, such as an approaching predator, and to participate in communal exchanges like those concerning territories or mating. On the other hand, although it is physically linked to the auditory system, the vestibular system is not involved in hearing. Instead, an animal's vestibular system detects its own movement, both linear and angular acceleration and deceleration, and balance.

Sound

Auditory stimuli are sound waves, which are mechanical, pressure waves that move through a medium, such as air or water. There are no sound waves in a vacuum since there are no air molecules to move in waves. The speed of sound waves differs, based on altitude, temperature, and medium, but at sea level and a temperature of 20º C (68º F), sound waves travel in the air at about 343 meters per second.

As is true for all waves, there are four main characteristics of a sound wave: frequency, wavelength, period, and amplitude. Frequency is the number of waves per unit of time, and in sound is heard as pitch. High-frequency (≥15.000Hz) sounds are higher-pitched (short wavelength) than low-frequency (long wavelengths; ≤100Hz) sounds. Frequency is measured in cycles per second, and for sound, the most commonly used unit is hertz (Hz), or cycles per second. Most humans can perceive sounds with

frequencies between 30 and 20,000 Hz. Women are typically better at hearing high frequencies, but everyone's ability to hear high frequencies decreases with age. Dogs detect up to about 40,000 Hz; cats, 60,000 Hz; bats, 100,000 Hz; and dolphins 150,000 Hz, and American shad (**Alosa sapidissima**), a fish, can hear 180,000 Hz. Those frequencies above the human range are called **ultrasound**.

Amplitude, or the dimension of a wave from peak to trough, in sound is heard as volume and is illustrated in [Figure 8.21](). The sound waves of louder sounds have greater amplitude than those of softer sounds. For sound, volume is measured in decibels (dB). The softest sound that a human can hear is the zero point. Humans speak normally at 60 decibels.
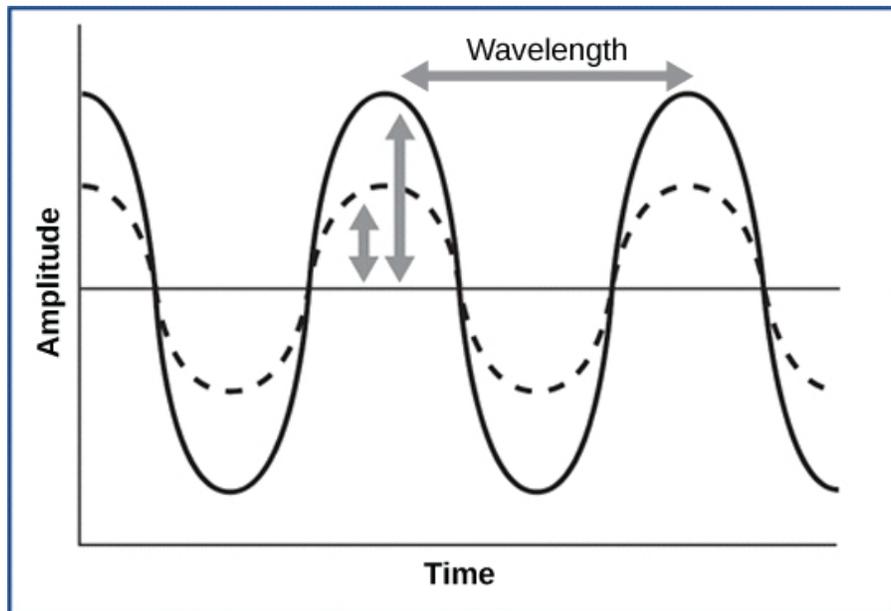

Figure 8.21. For sound waves, wavelength corresponds to pitch. Amplitude of the wave corresponds to volume. The sound wave shown with a dashed line is softer in volume than the sound wave shown with a solid line. (credit: NIH)

Reception of sound

In mammals, sound waves are collected by the external, cartilaginous part of the ear called the **pinna**, then travel through the auditory canal and cause vibration of the thin diaphragm called the **tympanum** or ear drum, the innermost part of the **outer ear** (illustrated in [Figure 8.22]()). Interior to the tympanum is the **middle ear**. The middle ear holds three small bones called the **ossicles,** which transfer energy from the moving tympanum to the inner ear. The three ossicles are the **malleus**(also known as the hammer), the **incus** (the anvil), and **stapes** (the stirrup). The aptly named stapes looks very much like a stirrup. The three ossicles are unique to mammals, and each plays a role in hearing. The malleus attaches at three points to the interior surface of the tympanic membrane. The incus attaches the malleus to the stapes. In humans, the stapes is not long enough to reach the tympanum.

If we did not have the malleus and the incus, then the vibrations of the tympanum would never reach the inner ear. These bones also function to collect force and amplify sounds. The ear ossicles are homologous to bones in a fish mouth: the bones that support gills in fish are thought to be adapted for use in the vertebrate ear over evolutionary time. Many animals (frogs, reptiles, and birds, for example) use the stapes of the middle ear to transmit vibrations to the middle ear.
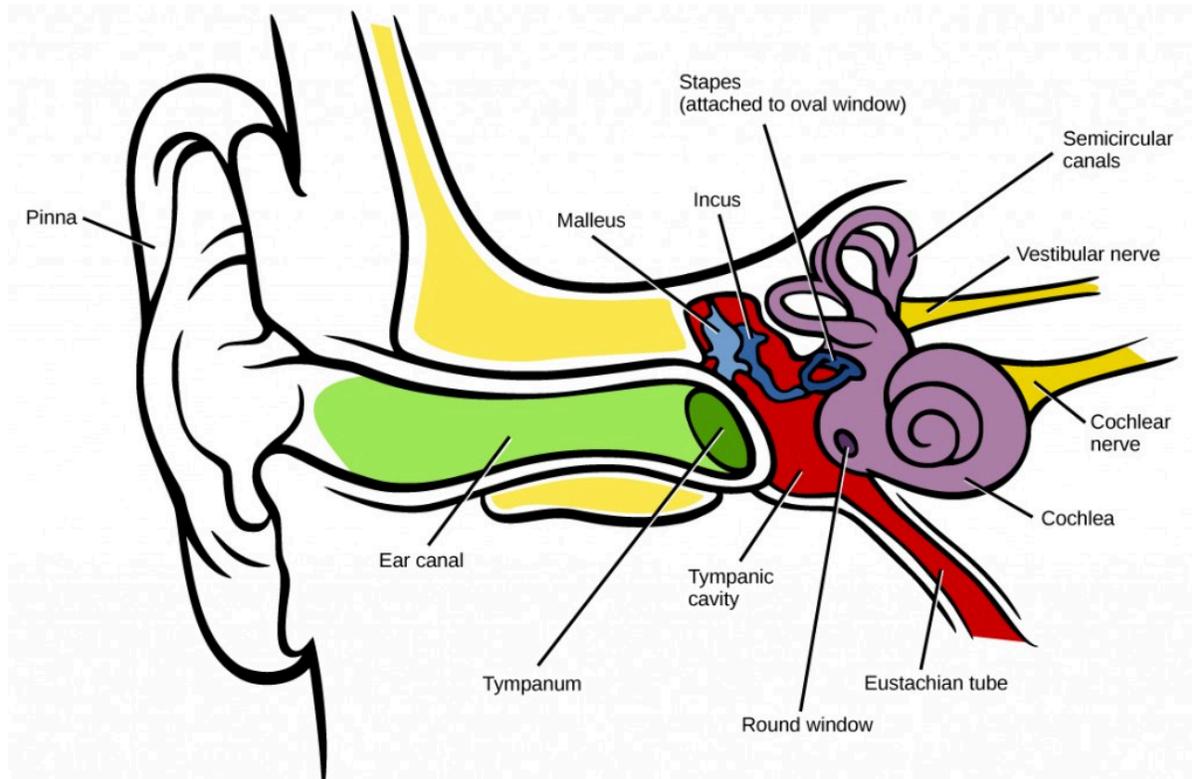

Figure 8.22. Sound travels through the outer ear to the middle ear, which is bounded on its exterior by the tympanic membrane. The middle ear contains three bones called ossicles that transfer the sound wave to the oval window, the exterior boundary of the inner ear. The organ of Corti, which is the organ of sound transduction, lies inside the cochlea. (credit: modification of work by Lars Chittka, Axel Brockmann)

Transduction of Sound

Vibrating objects, such as vocal cords, create sound waves or pressure waves in the air. When these pressure waves reach the ear, the ear transduces this mechanical stimulus (pressure wave) into a nerve impulse (electrical signal) that the brain perceives as sound. The pressure waves strike the tympanum, causing it to vibrate. The mechanical energy from the moving tympanum transmits the vibrations to the three bones of the middle ear. The stapes transmits the vibrations to a thin diaphragm called the **oval window**, which is the outermost structure of the **inner ear**. The structures of the inner ear are found in the **labyrinth**, a bony, hollow structure that is the most interior portion of the ear. Here, the energy from the sound wave is transferred from the stapes through the flexible

oval window and to the fluid of the cochlea. The vibrations of the oval window create pressure waves in the fluid (perilymph) inside the cochlea. The **cochlea** is a whorled structure, like the shell of a snail, and it contains receptors for transduction of the mechanical wave into an electrical signal (as illustrated in [Figure 8.23](#)). Inside the cochlea, the **basilar membrane**is a mechanical analyzer that runs the length of the cochlea, curling toward the cochlea's center.

The mechanical properties of the basilar membrane change along its length, such that it is thicker, tauter, and narrower at the outside of the whorl (where the cochlea is largest), and thinner, floppier, and broader toward the apex, or center, of the whorl (where the cochlea is smallest). Different regions of the basilar membrane vibrate according to the frequency of the sound wave conducted through the fluid in the cochlea. For these reasons, the fluid-filled cochlea detects different wave frequencies (pitches) at different regions of the membrane. When the sound waves in the cochlear fluid contact the basilar membrane, it flexes back and forth in a wave-like fashion. Above the basilar membrane is the **tectorial membrane**.
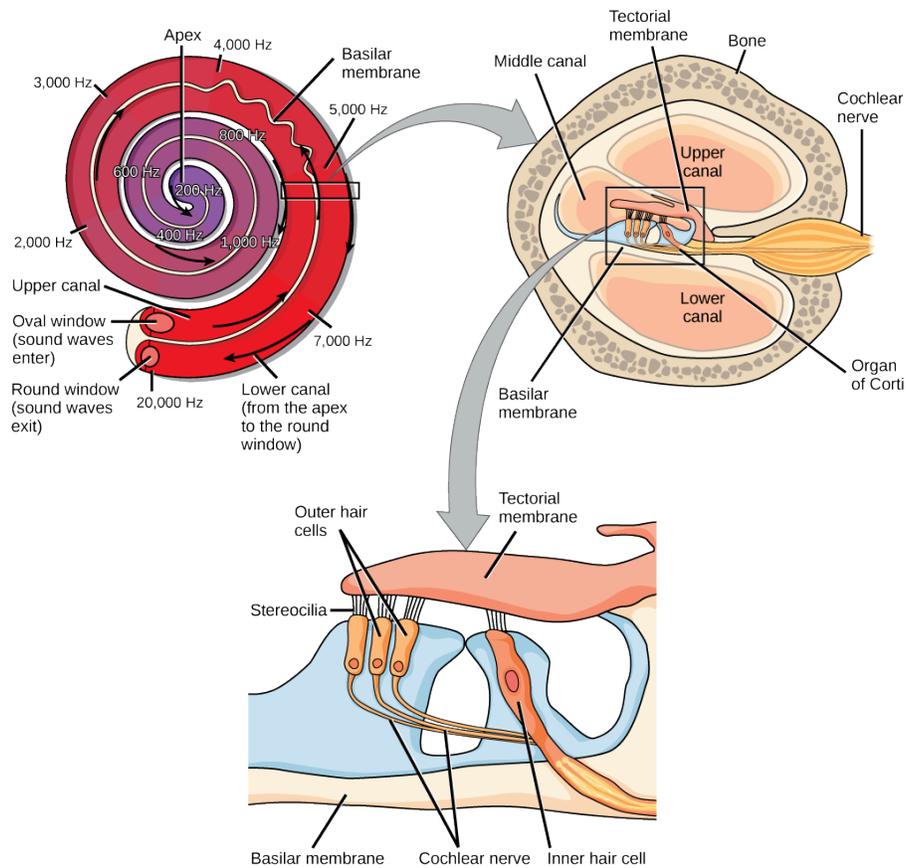
Figure 8.23. In the human ear, sound waves cause the stapes to press against the oval window. Vibrations travel up the fluid-filled interior of the cochlea. The basilar membrane that lines the cochlea gets continuously thinner toward the apex of the cochlea. Different thicknesses of membrane vibrate in response to different frequencies of sound. Sound waves then exit through the round window. In the cross section of the cochlea (top right figure), note that in addition to the upper canal and lower canal, the cochlea also has a middle canal. The organ of Corti (bottom image) is the site of sound transduction. Movement of stereocilia on hair cells results in an action potential that travels along the auditory nerve.

Question 8.17

Cochlear implants can restore hearing in people who have a nonfunctional cochlear. The implant consists of a microphone that picks up sound. A speech processor selects sounds in the range of human speech, and a transmitter converts these sounds to electrical impulses, which are then sent to the auditory nerve. Which of the following types of hearing loss would not be restored by a cochlear implant?

a. Hearing loss resulting from absence or loss of hair cells in the organ of Corti.
b. Hearing loss resulting from an abnormal auditory nerve.
c. Hearing loss resulting from fracture of the cochlea.
d. Hearing loss resulting from damage to bones of the middle ear.

The site of transduction is in the **organ of Corti** (spiral organ). It is composed of hair cells held in place above the basilar membrane like flowers projecting up from soil, with their exposed short, hair-like **stereocilia** contacting or embedded in the tectorial membrane above them. The inner hair cells are the primary auditory receptors and exist in a single row, numbering approximately 3,500. The stereocilia from inner hair cells extend into small dimples on the tectorial membrane's lower surface. The outer hair cells are arranged in three or four rows. They number approximately 12,000, and they function to fine tune incoming sound waves. The longer stereocilia that project from the outer hair cells actually attach to the tectorial membrane. All of the stereocilia are mechanoreceptors, and when bent by vibrations they respond by opening a gated ion channel (refer to Figure 8.22). As a result, the hair cell membrane is depolarized, and a signal is transmitted to the chochlear nerve. Intensity (volume) of sound is determined by how many hair cells at a particular location are stimulated.

The hair cells are arranged on the basilar membrane in an orderly way. The basilar membrane vibrates in different regions, according to the frequency of the sound waves impinging on it. Likewise, the hair cells that lay above it are most sensitive to a specific frequency of sound waves. Hair cells can respond to a small range of similar frequencies, but they require stimulation of greater intensity to fire at frequencies outside of their optimal range. The difference in response frequency between adjacent inner hair cells is about 0.2 percent. Compare that to adjacent piano strings, which are about six percent different. Place theory, which is the model for how biologists think pitch detection works in the human ear, states that high frequency sounds selectively vibrate the basilar membrane of the inner ear near the entrance port (the oval window). Lower frequencies travel farther along the membrane before causing appreciable excitation of the membrane. The basic pitch-determining mechanism is based on the location along the membrane where the hair cells are stimulated. The place theory is the first step toward an understanding of pitch perception. Considering the extreme pitch sensitivity of the human ear, it is thought that there must be some auditory "sharpening" mechanism to enhance the pitch resolution.

When sound waves produce fluid waves inside the cochlea, the basilar membrane flexes, bending the stereocilia that attach to the tectorial membrane. Their bending results in action potentials in the hair cells, and auditory information travels along the neural endings of the bipolar neurons of the hair cells (collectively, the auditory nerve) to the brain. When the hairs bend, they release an excitatory neurotransmitter at a synapse with a sensory neuron, which then conducts action potentials to the central nervous system. The cochlear branch of the vestibulocochlear cranial nerve

sends information on hearing. The auditory system is very refined, and there is some modulation or "sharpening" built in. The brain can send signals back to the cochlea, resulting in a change of length in the outer hair cells, sharpening or dampening the hair cells' response to certain frequencies.

Watch an [animation](#) of sound entering the outer ear, moving through the ear structure, stimulating cochlear nerve impulses, and eventually sending signals to the temporal lobe.

Higher processing

The inner hair cells are most important for conveying auditory information to the brain. About 90 percent of the afferent neurons carry information from inner hair cells, with each hair cell synapsing with 10 or so neurons. Outer hair cells connect to only 10 percent of the afferent neurons, and each afferent neuron innervates many hair cells. The afferent, bipolar neurons that convey auditory information travel from the cochlea to the medulla, through the pons and midbrain in the brainstem, finally reaching the primary auditory cortex in the temporal lobe.

Vestibular information

The stimuli associated with the vestibular system are linear acceleration (gravity) and angular acceleration and deceleration. Gravity, acceleration, and deceleration are detected by evaluating the inertia on receptive cells in the vestibular system. Gravity is detected through head position. Angular acceleration and deceleration are expressed through turning or tilting of the head.

The vestibular system has some similarities with the auditory system. It utilizes hair cells just like the auditory system, but it excites them in different ways. There are five vestibular receptor organs in the inner ear: the utricle, the saccule, and three semicircular canals. Together, they make up what's known as the vestibular labyrinth that is shown in [Figure 8.24](#). The utricle and saccule respond to acceleration in a straight line, such as gravity. The roughly 30,000 hair cells in the utricle and 16,000 hair cells in the saccule lie below a gelatinous layer, with their stereocilia projecting into the gelatin. Embedded in this gelatin are calcium carbonate crystals–like tiny rocks. When the head is tilted, the crystals continue to be pulled straight down by gravity, but the new angle of the head causes the gelatin to shift, thereby bending the stereocilia. The bending of the stereocilia stimulates the neurons,

and they signal to the brain that the head is tilted, allowing the maintenance of balance. It is the vestibular branch of the vestibulocochlear cranial nerve that deals with balance.
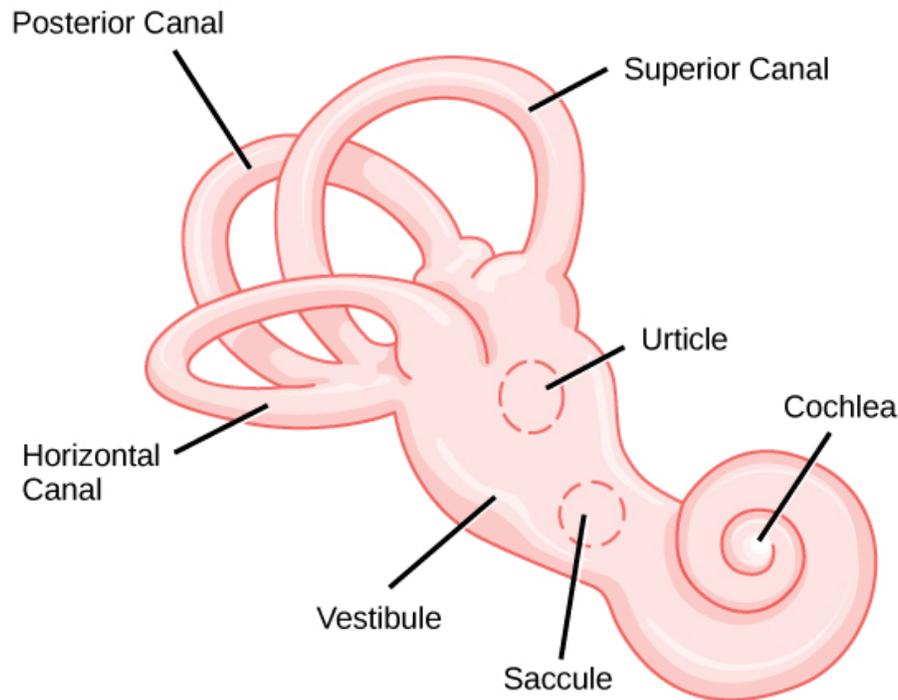

Figure 8.24. The structure of the vestibular labyrinth is shown. (credit: modification of work by NIH)

The fluid-filled **semicircular canals** are tubular loops set at oblique angles. They are arranged in three spatial planes. The base of each canal has a swelling that contains a cluster of hair cells. The hairs project into a gelatinous cap called the cupula and monitor angular acceleration and deceleration from rotation. They would be stimulated by driving your car around a corner, turning your head, or falling forward. One canal lies horizontally, while the other two lie at about 45 degree angles to the horizontal axis, as illustrated in Figure 8.24. When the brain processes input from all three canals together, it can detect angular acceleration or deceleration in three dimensions. When the head turns, the fluid in the canals shifts, thereby bending stereocilia and sending signals to the brain. Upon cessation accelerating or decelerating–or just moving–the movement of the fluid within the canals slows or stops. For example, imagine holding a glass of water. When moving forward, water may splash backwards onto the hand, and when motion has stopped, water may splash forward onto the fingers. While in motion, the water settles in the glass and does not splash. Note that the canals are not sensitive to velocity itself, but to changes in velocity, so moving forward at 60mph with your eyes closed would not give the sensation of movement, but suddenly accelerating or braking would stimulate the receptors.

Hair cells from the utricle, saccule, and semicircular canals also communicate through bipolar

neurons to the cochlear nucleus in the medulla. Cochlear neurons send descending projections to the spinal cord and ascending projections to the pons, thalamus, and cerebellum. Connections to the cerebellum are important for coordinated movements. There are also projections to the temporal cortex, which account for feelings of dizziness; projections to autonomic nervous system areas in the brainstem, which account for motion sickness; and projections to the primary somatosensory cortex, which monitors subjective measurements of the external world and self-movement. People with lesions in the vestibular area of the somatosensory cortex see vertical objects in the world as being tilted. Finally, the vestibular signals project to certain optic muscles to coordinate eye and head movements.

Click through this interactive tutorial to review the parts of the ear and how they function to process sound.

Question 8.18

Auditory hair cells are indirectly anchored to the _____.
a. basilar membrane
b. oval window
c. tectorial membrane
d. ossicles

Question 8.19

Which of the following are found both in the auditory system and the vestibular system?
a. basilar membrane
b. hair cells
c. semicircular canals
d. ossicles

Question 8.20

How would a rise in altitude likely affect the speed of a sound transmitted through air?

Question 8.21

How might being in a place with less gravity than Earth has (such as Earth's moon) affect vestibular sensation, and why?

Vision

Vision is the ability to detect light patterns from the outside environment and interpret them into images. Animals are bombarded with sensory information, and the sheer volume of visual information can be problematic. Fortunately, the visual systems of species have evolved to attend to the most-important stimuli. The importance of vision to humans is further substantiated by the fact that about one-third of the human cerebral cortex is dedicated to analyzing and perceiving visual information.

Light

As with auditory stimuli, light travels in waves. The compression waves that compose sound must travel in a medium—a gas, a liquid, or a solid. In contrast, light is composed of electromagnetic waves and needs no medium; light can travel in a vacuum ([Figure 8.25](#)). The behavior of light can be discussed in terms of the behavior of waves and also in terms of the behavior of the fundamental unit of light—a packet of electromagnetic radiation called a photon. A glance at the electromagnetic spectrum shows that visible light for humans is just a small slice of the entire spectrum, which includes radiation that we cannot see as light because it is below the frequency of visible red light and above the frequency of visible violet light.

Certain variables are important when discussing perception of light. Wavelength (which varies inversely with frequency) manifests itself as hue. Light at the red end of the visible spectrum has longer wavelengths (and is lower frequency), while light at the violet end has shorter wavelengths (and is higher frequency). The wavelength of light is expressed in nanometers (nm); one nanometer is one billionth of a meter. Humans perceive light that ranges between approximately 380 nm and 740 nm. Some other animals, though, can detect wavelengths outside of the human range. For example, bees see near-ultraviolet light in order to locate nectar guides on flowers, and some non-avian reptiles sense infrared light (heat that prey gives off).
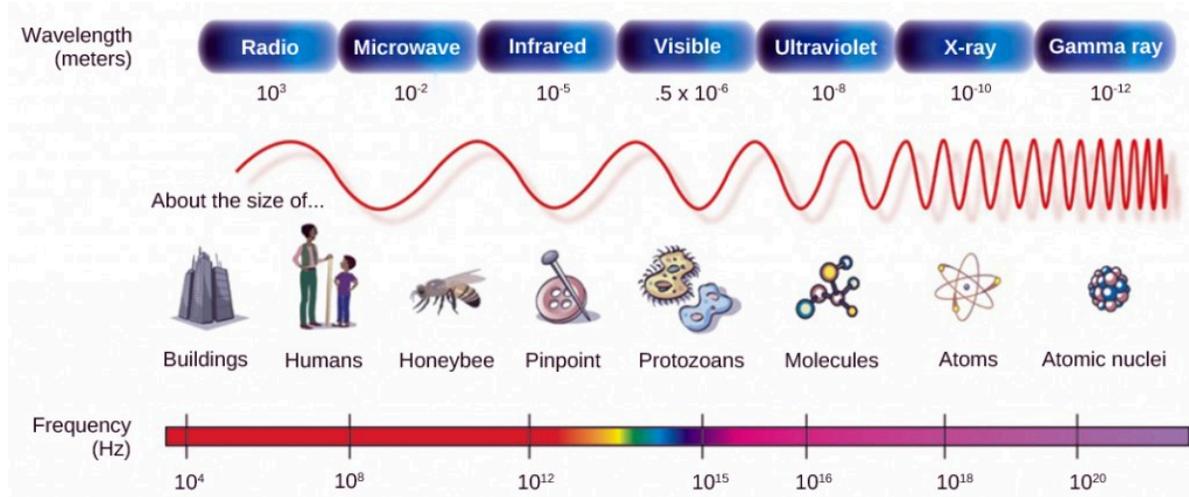
Figure 8.25. In the electromagnetic spectrum, visible light lies between 380 nm and 740 nm. (credit: modification of work by NASA)

Wave amplitude is perceived as luminous intensity, or brightness. The standard unit of intensity of light is the *candela*, which is approximately the luminous intensity of a one common candle.

Light waves travel 299,792 km per second in a vacuum, (and somewhat slower in various media such as air and water), and those waves arrive at the eye as long (red), medium (green), and short (blue) waves. What is termed "white light" is light that is perceived as white by the human eye. This effect is produced by light that stimulates equally the color receptors in the human eye. The apparent color of an object is the color (or colors) that the object reflects. Thus a red object reflects the red wavelengths in mixed (white) light and absorbs all other wavelengths of light.

Anatomy of the eye

The photoreceptive cells of the eye, where transduction of light to nervous impulses occurs, are located in the *retina* (Figure 8.26) on the inner surface of the back of the eye. But light does not impinge on the retina unaltered. It passes through other layers that process it so that it can be interpreted by the retina (Figure 8.25b). The **cornea,** the front transparent layer of the eye, and the crystalline **lens**, a transparent convex structure behind the cornea, both refract (bend) light to focus the image on the retina. The **iris**, which is conspicuous as the colored part of the eye, is a circular muscular ring lying between the lens and cornea that regulates the amount of light entering the eye. In conditions of high ambient light, the iris contracts, reducing the size of the pupil at its center. In conditions of low light, the iris relaxes and the pupil enlarges.
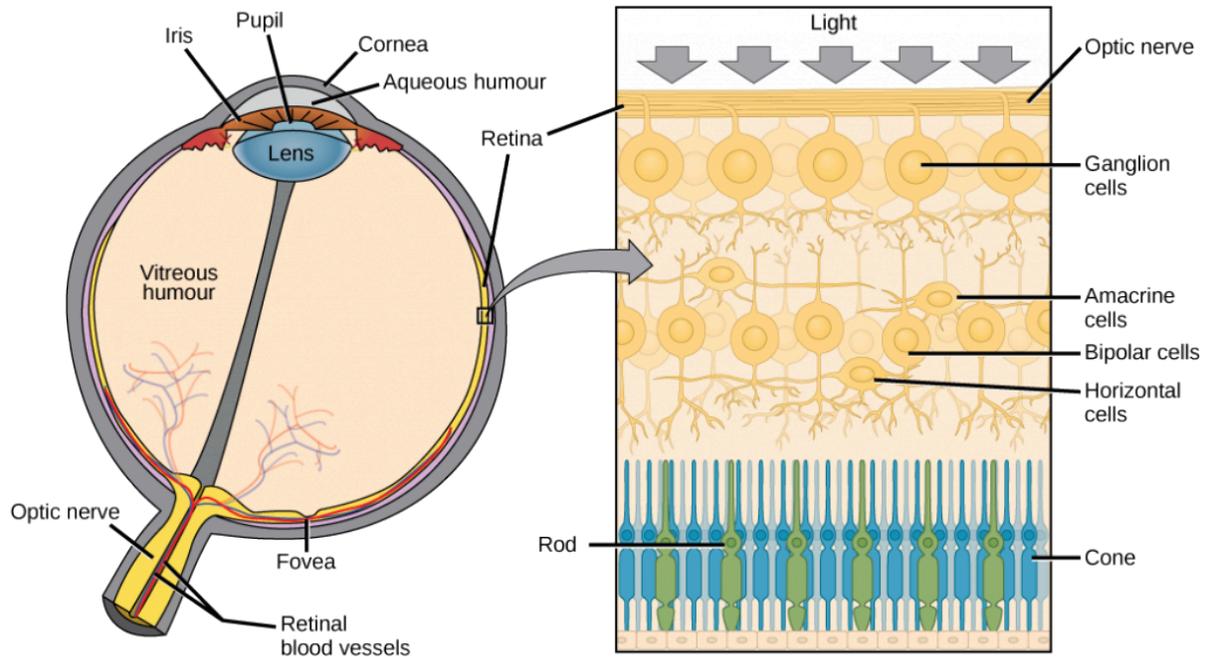
Figure 8.26. (a) The human eye is shown in cross section. (b) A blowup shows the layers of the retina.

Question 8.22

Which of the following statements about the human eye is false?

a. Rods detect color, while cones detect only shades of gray.
b. When light enters the retina, it passes the ganglion cells and bipolar cells before reaching photoreceptors at the rear of the eye.
c. The iris adjusts the amount of light coming into the eye.
d. The cornea is a protective layer on the front of the eye.

There are two types of photoreceptors in the retina: **rods** and **cones**, named for their general appearance as illustrated in Figure 8.27. Rods are strongly photosensitive and are located in the outer edges of the retina. They detect dim light and are used primarily for peripheral and nighttime vision. Cones are weakly photosensitive and are located near the center of the retina. They respond to bright light, and their primary role is in daytime, color vision.
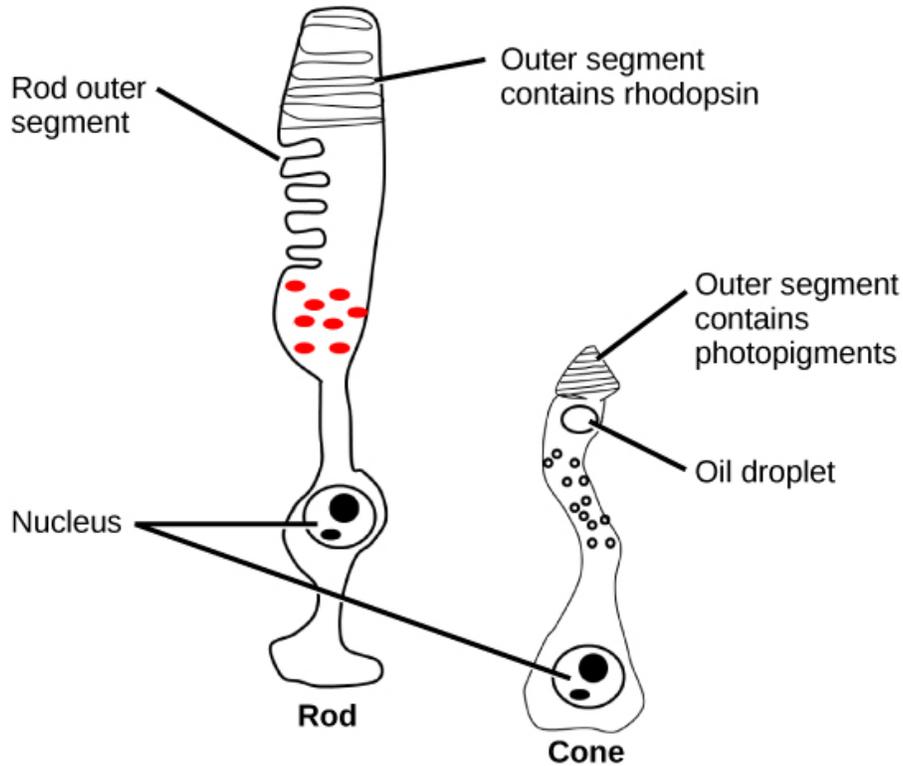
Figure 8.27. Rods and cones are photoreceptors in the retina. Rods respond in low light and can detect only shades of gray. Cones respond in intense light and are responsible for color vision. (credit: modification of work by Piotr Sliwa)

The **fovea** is the region in the center back of the eye that is responsible for acute vision. The fovea has a high density of cones. When you bring your gaze to an object to examine it intently in bright light, the eyes orient so that the object's image falls on the fovea. However, when looking at a star in the night sky or other object in dim light, the object can be better viewed by the peripheral vision because it is the rods at the edges of the retina, rather than the cones at the center, that operate better in low light. In humans, cones far outnumber rods in the fovea.

Review the anatomical structure of the eye, clicking on each part to practice identification.

Transduction of light

The rods and cones are the site of transduction of light to a neural signal. Both rods and cones contain photopigments. In vertebrates, the main photopigment, **rhodopsin**, has two main parts: an opsin, which is a membrane protein (in the form of a cluster of α-helices that span the membrane), and retinal—a molecule that absorbs light. When light hits a photoreceptor, it causes a shape change in the retinal, altering its structure from a bent (*cis*) form of the molecule to its linear (*trans*) isomer. This isomerization of retinal activates the rhodopsin, starting a cascade of events that ends with the closing of Na^+ channels in the membrane of the photoreceptor. Thus, unlike most other sensory neurons (which become depolarized by exposure to a stimulus) visual receptors become hyperpolarized and thus driven away from threshold (Figure 8.28).
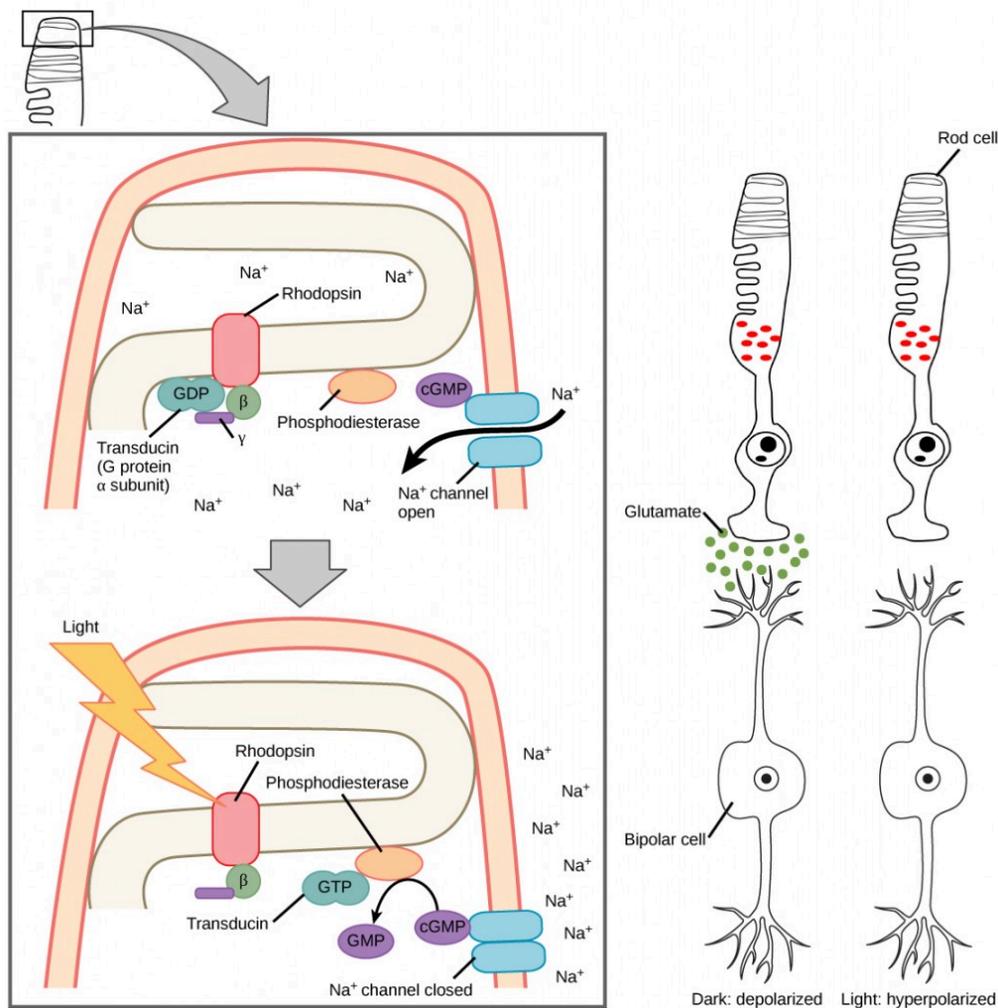

Figure 8.28. When light strikes rhodopsin, the G-protein transducin is activated, which in turn activates phosphodiesterase. Phosphodiesterase converts cGMP to GMP, thereby closing sodium channels. As a result, the membrane becomes hyperpolarized. The hyperpolarized membrane does not release glutamate to the bipolar cell.

Higher processing

The myelinated axons of ganglion cells make up the optic nerves. Within the nerves, different axons carry different qualities of the visual signal. Some axons constitute the magnocellular (big cell) pathway, which carries information about form, movement, depth, and differences in brightness. Other axons constitute the parvocellular (small cell) pathway, which carries information on color and fine detail. Some visual information projects directly back into the brain, while other information crosses to the opposite side of the brain. This crossing of optical pathways produces the distinctive optic chiasma (Greek, for "crossing") found at the base of the brain and allows us to coordinate information from both eyes.

Once in the brain, visual information is processed in several places, and its routes reflect the complexity and importance of visual information to humans and other animals. One route takes the signals to the thalamus, which serves as the routing station for all incoming sensory impulses except olfaction. In the thalamus, the magnocellular and parvocellular distinctions remain intact, and there are different layers of the thalamus dedicated to each. When visual signals leave the thalamus, they travel to the primary visual cortex at the rear of the brain. From the visual cortex, the visual signals travel in two directions. One stream that projects to the parietal lobe, in the side of the brain, carries magnocellular ("where") information. A second stream projects to the temporal lobe and carries both magnocellular ("where") and parvocellular ("what") information.

Another important visual route is a pathway from the retina to the **superior colliculus** in the midbrain, where eye movements are coordinated and integrated with auditory information. Finally, there is the pathway from the retina to the **suprachiasmatic nucleus** (SCN) of the hypothalamus. The SCN is a cluster of cells that is considered to be the body's internal clock, which controls our **circadian** (day-long) cycle. The SCN sends information to the pineal gland, which is important in sleep/wake patterns and annual cycles.

View this [interactive presentation](#) to review what you have learned about how vision functions.

Question 8.23

Why is it easier to see images at night using peripheral, rather than the central, vision?

a. Cones are denser in the periphery of the retina.
b. Bipolar cells are denser in the periphery of the retina.
c. Rods are denser in the periphery of the retina.
d. The optic nerve exits at the periphery of the retina.

Question 8.24

How could the pineal gland, the brain structure that plays a role in annual cycles, use visual information from the suprachiasmatic nucleus of the hypothalamus?

Question 8.25

What sensory information do animals use that human cannot? Provide three different examples and briefly describe each.

CHAPTER 9 – MAINTAINING INTERNAL BALANCE

9.1. Explain various external cues and their relevance for animal internal balance;

9.2. Describe general cell signaling strategies and explain types of hormones and their signaling pathways in endocrine communication; and,

9.3. Use different hormone examples (molting, digestion and circadian rhythm-related) to explain their role in maintaining homeostasis.

The content of this chapter was adapted from the *Concepts of Biology–1st Canadian Edition* open textbook by Charles Molnar and Jane Gair and *Anatomy and Physiology* open textbook.

9.1 – Interpreting External Cues

9.1. Explain various external cues and their relevance for animal internal balance.

An animal's endocrine system controls body processes through the production, secretion, and regulation of hormones, which serve as chemical "messengers" functioning in cellular and organ activity and, ultimately, maintaining the body's homeostasis. The endocrine system plays a role in growth, metabolism, and sexual development. In humans, common endocrine system diseases include thyroid disease and diabetes mellitus. In organisms that undergo metamorphosis, the process is controlled by the endocrine system. The transformation from tadpole to frog, for example, is complex and nuanced to adapt to specific environments and ecological circumstances (Figure 9.1). The endocrine system also plays an important role in the timing of migration in animals that live in different regions in summer than in winter.

Figure 9.1. The process of amphibian metamorphosis, as seen in the tadpole-to-frog stages shown here, is driven by hormones. (credit "tadpole": modification of work by Brian Gratwicke).

Here is an article summarizing animal migration. This article may help you with the Question 9.1 below and will provide an introduction how important the timing of environmental cues is important for animal life cycle.

Question 9.1

How does interpretation of external environment impact animal survival or internal environment? Provide at least one example.

9.2 – Signaling Pathways, Hormones and Endocrine System

9.2. Describe general cell signaling strategies and explain types of hormones and their signaling pathways in endocrine communication.

Chemical signaling

There are two kinds of communication in the world of living cells. Communication between cells is called intercellular signaling, and communication within a cell is called intracellular signaling. An easy way to remember the distinction is by understanding the Latin origin of the prefixes: inter- means "between" (for example, intersecting lines are those that cross each other) and intra- means "inside" (like intravenous).

Chemical signals are released by signaling cells in the form of small, usually volatile or soluble molecules called ligands. A ligand is a molecule that binds another specific molecule, in some cases, delivering a signal in the process. Ligands can thus be thought of as signaling molecules. Ligands interact with proteins in target cells, which are cells that are affected by chemical signals; these proteins are also called receptors. Ligands and receptors exist in several varieties; however, a specific ligand will have a specific receptor that typically binds only that ligand.

Forms of Signaling

There are four categories of chemical signaling found in multicellular organisms: paracrine signaling, endocrine signaling, autocrine signaling, and direct signaling across gap junctions (Figure 9.2). The main difference between the different categories of signaling is the distance that the signal travels through the organism to reach the target cell. Not all cells are affected by the same signals.

Forms of Chemical Signaling	
Autocrine	A cell targets itself.
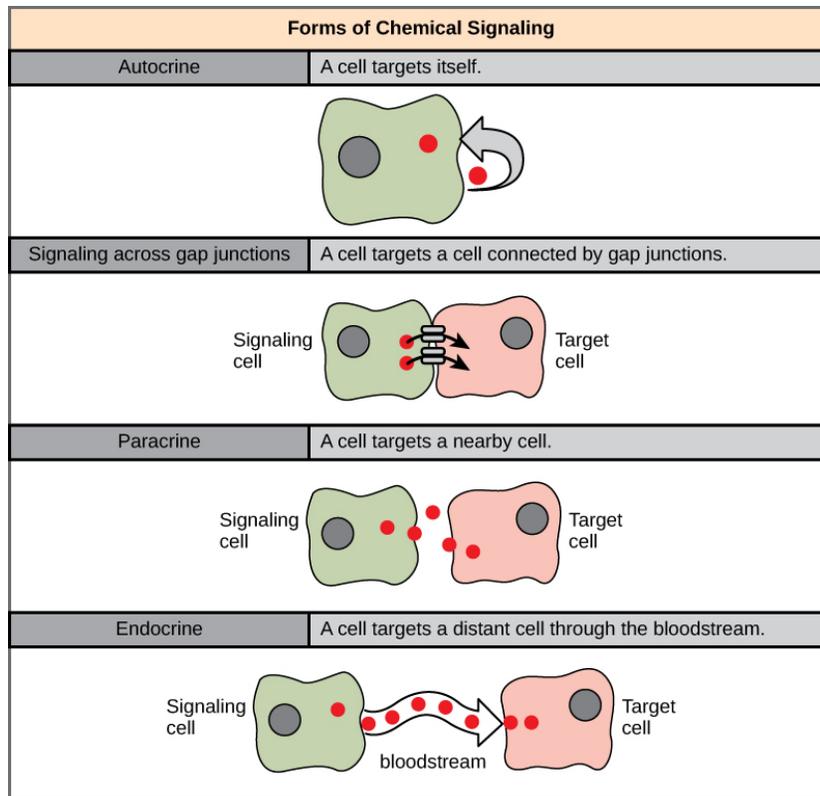
Figure 9.2. Forms of chemical signaling: autocrine, gap junctions, paracrine, and endocrine.

In chemical signaling, a cell may target itself (autocrine signaling), a cell connected by gap junctions, a nearby cell (paracrine signaling), or a distant cell (endocrine signaling). Paracrine signaling acts on nearby cells, endocrine signaling uses the circulatory system to transport ligands, and autocrine signaling acts on the signaling cell. Signaling via gap junctions involves signaling molecules moving directly between adjacent cells.

Paracrine signaling

Signals that act locally between cells that are close together are called paracrine signals. Paracrine signals move by diffusion through the extracellular matrix. These types of signals usually elicit quick responses that last only a short amount of time. In order to keep the response localized, paracrine ligand molecules are normally quickly degraded by enzymes or removed by neighboring cells. Removing the signals will reestablish the concentration gradient for the signal, allowing them to quickly diffuse through the intracellular space if released again.

One example of paracrine signaling is the transfer of signals across synapses between nerve cells.

A nerve cell consists of a cell body, several short, branched extensions called dendrites that receive stimuli, and a long extension called an axon, which transmits signals to other nerve cells or muscle cells. The junction between nerve cells where signal transmission occurs is called a synapse. A synaptic signal is a chemical signal that travels between nerve cells. Signals within the nerve cells are propagated by fast-moving electrical impulses. When these impulses reach the end of the axon, the signal continues on to a dendrite of the next cell by the release of chemical ligands called neurotransmitters by the presynaptic cell (the cell emitting the signal). The neurotransmitters are transported across the very small distances between nerve cells, which are called chemical synapses (Figure 9.3). The small distance between nerve cells allows the signal to travel quickly; this enables an immediate response, such as, Take your hand off the stove!

When the neurotransmitter binds the receptor on the surface of the postsynaptic cell, the electrochemical potential of the target cell changes, and the next electrical impulse is launched. The neurotransmitters that are released into the chemical synapse are degraded quickly or get reabsorbed by the presynaptic cell so that the recipient nerve cell can recover quickly and be prepared to respond rapidly to the next synaptic signal.
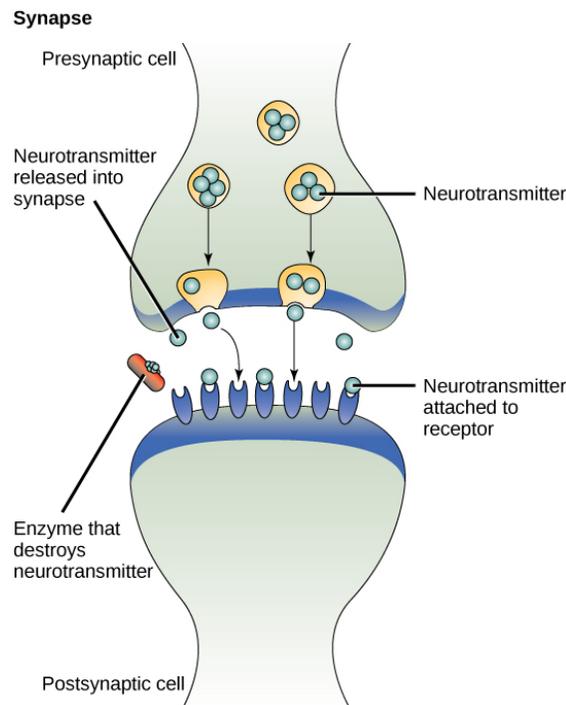
Figure 9.3. Synapse showing neurotransmitter release.

The distance between the presynaptic cell and the postsynaptic cell–called the synaptic gap–is very small and allows for rapid diffusion of the neurotransmitter. Enzymes in the synapatic cleft degrade some types of neurotransmitters to terminate the signal.

Endocrine signaling

Signals from distant cells are called endocrine signals, and they originate from endocrine cells. (In the body, many endocrine cells are located in endocrine glands, such as the thyroid gland, the hypothalamus, and the pituitary gland.) These types of signals usually produce a slower response but have a longer-lasting effect. The ligands released in endocrine signaling are called hormones, signaling molecules that are produced in one part of the body but affect other body regions some distance away.

Hormones travel the large distances between endocrine cells and their target cells via the bloodstream, which is a relatively slow way to move throughout the body. Because of their form of transport, hormones get diluted and are present in low concentrations when they act on their target cells. This is different from paracrine signaling, in which local concentrations of ligands can be very high.

Autocrine signaling

Autocrine signals are produced by signaling cells that can also bind to the ligand that is released. This means the signaling cell and the target cell can be the same or a similar cell (the prefix *auto-* means self, a reminder that the signaling cell sends a signal to itself). This type of signaling often occurs during the early development of an organism to ensure that cells develop into the correct tissues and take on the proper function. Autocrine signaling also regulates pain sensation and inflammatory responses. Further, if a cell is infected with a virus, the cell can signal itself to undergo programmed cell death, killing the virus in the process. In some cases, neighboring cells of the same type are also influenced by the released ligand. In embryological development, this process of stimulating a group of neighboring cells may help to direct the differentiation of identical cells into the same cell type, thus ensuring the proper developmental outcome.

Direct signaling across gap junctions

Gap junctions in animals and plasmodesmata in plants are connections between the plasma membranes of neighboring cells. These water-filled channels allow small signaling molecules, called intracellular mediators, to diffuse between the two cells. Small molecules, such as calcium ions ($Ca2+$), are able to move between cells, but large molecules like proteins and DNA cannot fit through the channels. The specificity of the channels ensures that the cells remain independent but can quickly and easily transmit signals. The transfer of signaling molecules communicates the current state of the cell that is directly next to the target cell; this allows a group of cells to coordinate

their response to a signal that only one of them may have received. In plants, plasmodesmata are ubiquitous, making the entire plant into a giant, communication network.

Types of receptors

Receptors are protein molecules in the target cell or on its surface that bind ligand. There are two types of receptors, internal receptors and cell-surface receptors.

Internal receptors

Internal receptors, also known as intracellular or cytoplasmic receptors, are found in the cytoplasm of the cell and respond to hydrophobic ligand molecules that are able to travel across the plasma membrane. Once inside the cell, many of these molecules bind to proteins that act as regulators of mRNA synthesis (transcription) to mediate gene expression. Gene expression is the cellular process of transforming the information in a cell's DNA into a sequence of amino acids, which ultimately forms a protein. When the ligand binds to the internal receptor, a conformational change is triggered that exposes a DNA-binding site on the protein. The ligand-receptor complex moves into the nucleus, then binds to specific regulatory regions of the chromosomal DNA and promotes the initiation of transcription ([Figure 9.4](#)). Transcription is the process of copying the information in a cells DNA into a special form of RNA called messenger RNA (mRNA); the cell uses information in the mRNA (which moves out into the cytoplasm and associates with ribosomes) to link specific amino acids in the correct order, producing a protein. Internal receptors can directly influence gene expression without having to pass the signal on to other receptors or messengers.
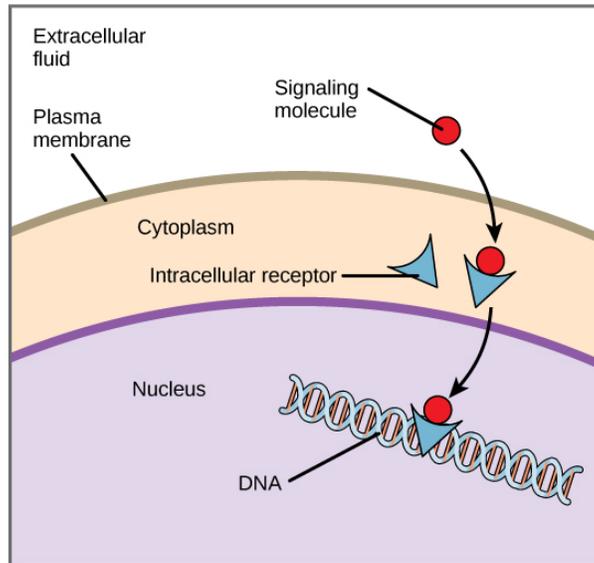
Figure 9.4. Hydrophobic signaling.

Hydrophobic signaling molecules typically diffuse across the plasma membrane and interact with intracellular receptors in the cytoplasm. Many intracellular receptors are transcription factors that interact with DNA in the nucleus and regulate gene expression.

Cell-surface receptors

Cell-surface receptors, also known as transmembrane receptors, are cell surface, membrane-anchored (integral) proteins that bind to external ligand molecules. This type of receptor spans the plasma membrane and performs signal transduction, in which an extracellular signal is converted into an intercellular signal. Ligands that interact with cell-surface receptors do not have to enter the cell that they affect. Cell-surface receptors are also called cell-specific proteins or markers because they are specific to individual cell types.

Because cell-surface receptor proteins are fundamental to normal cell functioning, it should come as no surprise that a malfunction in any one of these proteins could have severe consequences. Errors in the protein structures of certain receptor molecules have been shown to play a role in hypertension (high blood pressure), asthma, heart disease, and cancer.

Each cell-surface receptor has three main components: an external ligand-binding domain, a hydrophobic membrane-spanning region, and an intracellular domain inside the cell. The ligand-binding domain is also called the extracellular domain. The size and extent of each of these domains vary widely, depending on the type of receptor. Cell-surface receptors are involved in most of the

signaling in multicellular organisms. There are three general categories of cell-surface receptors: ion channel-linked receptors, G-protein-linked receptors, and enzyme-linked receptors.

Ion channel-linked receptors bind a ligand and open a channel through the membrane that allows specific ions to pass through. To form a channel, this type of cell-surface receptor has an extensive membrane-spanning region. In order to interact with the phospholipid fatty acid tails that form the center of the plasma membrane, many of the amino acids in the membrane-spanning region are hydrophobic in nature. Conversely, the amino acids that line the inside of the channel are hydrophilic to allow for the passage of water or ions. When a ligand binds to the extracellular region of the channel, there is a conformational change in the proteins structure that allows ions such as sodium, calcium, magnesium, and hydrogen to pass through ([Figure 9.5](#))
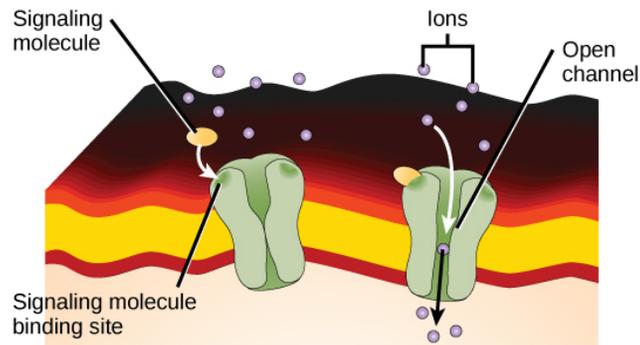

Figure 9.5. A closed gated ion channel. Gated ion channels form a pore through the plasma membrane that opens when the signaling molecule binds. The open pore then allows ions to flow into or out of the cell.

G-protein-linked receptors bind a ligand and activate a membrane protein called a G-protein. The activated G-protein then interacts with either an ion channel or an enzyme in the membrane ([Figure 9.6](#)). All G-protein-linked receptors have seven transmembrane domains, but each receptor has its own specific extracellular domain and G-protein-binding site.

Cell signaling using G-protein-linked receptors occurs as a cyclic series of events. Before the ligand binds, the inactive G-protein can bind to a newly revealed site on the receptor specific for its binding. Once the G-protein binds to the receptor, the resultant shape change activates the G-protein, which releases GDP and picks up GTP. The subunits of the G-protein then split into the α subunit and the $\beta\gamma$ subunit. One or both of these G-protein fragments may be able to activate other proteins as a result. After awhile, the GTP on the active α subunit of the G-protein is hydrolyzed to GDP and the $\beta\gamma$ subunit is deactivated. The subunits reassociate to form the inactive G-protein and the cycle begins anew.
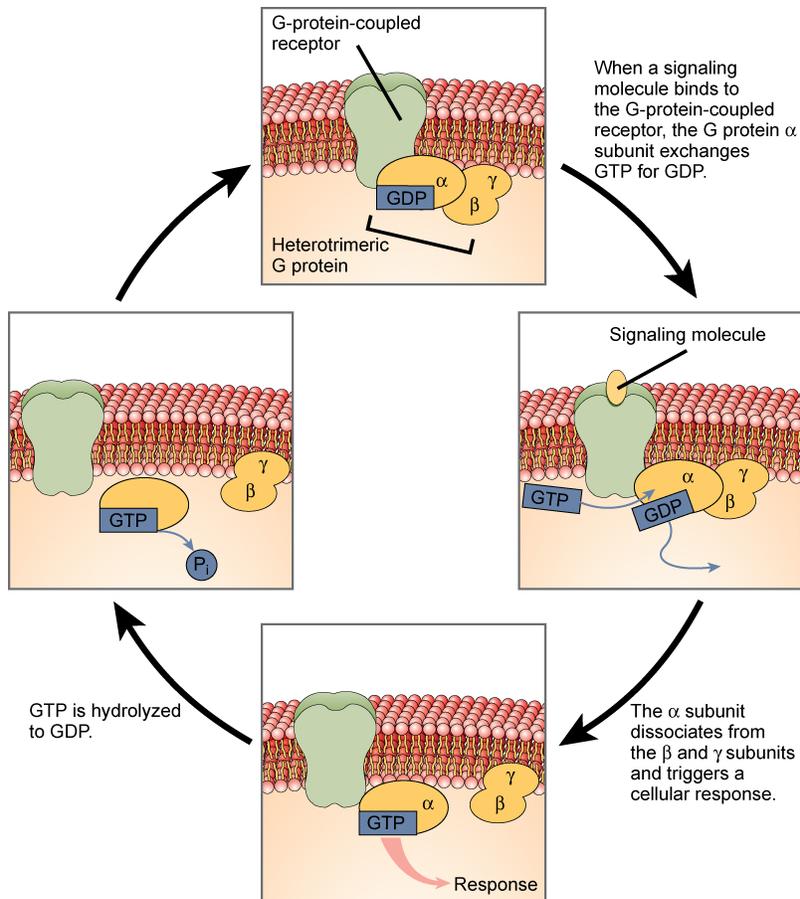
Figure 9.6. Heterotrimeric G proteins have three subunits: α, β, and γ. When a signaling molecule binds to a G-protein-coupled receptor in the plasma membrane, a GDP molecule associated with the α subunit is exchanged for GTP. The β and γ subunits dissociate from the α subunit, and a cellular response is triggered either by the α subunit or the dissociated βγ pair. Hydrolysis of GTP to GDP terminates the signal.

G-protein-linked receptors have been extensively studied and much has been learned about their roles in maintaining health. Bacteria that are pathogenic to humans can release poisons that interrupt specific G-protein-linked receptor function, leading to illnesses such as pertussis, botulism, and cholera. In cholera (Figure 9.7), for example, the water-borne bacterium *Vibrio cholerae* produces a toxin, choleragen, that binds to cells lining the small intestine. The toxin then enters these intestinal cells, where it modifies a G-protein that controls the opening of a chloride channel and causes it to remain continuously active, resulting in large losses of fluids from the body and potentially fatal dehydration as a result.
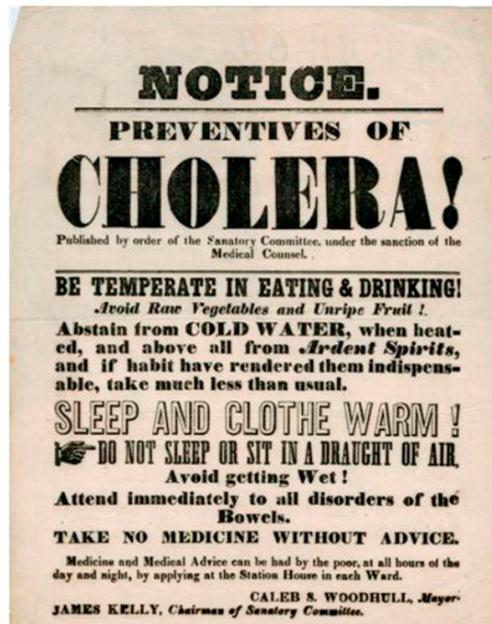
Figure 9.7. Transmitted primarily through contaminated drinking water, cholera is a major cause of death in the developing world and in areas where natural disasters interrupt the availability of clean water. The cholera bacterium, Vibrio cholerae, creates a toxin that modifies G-protein-mediated cell signaling pathways in the intestines. Modern sanitation eliminates the threat of cholera outbreaks, such as the one that swept through New York City in 1866. This poster from that era shows how, at that time, the way that the disease was transmitted was not understood. (credit: New York City Sanitary Commission)

Enzyme-linked receptors are cell-surface receptors with intracellular domains that are associated with an enzyme. In some cases, the intracellular domain of the receptor itself is an enzyme. Other enzyme-linked receptors have a small intracellular domain that interacts directly with an enzyme. The enzyme-linked receptors normally have large extracellular and intracellular domains, but the membrane-spanning region consists of a single alpha-helical region of the peptide strand. When a ligand binds to the extracellular domain, a signal is transferred through the membrane, activating the enzyme. Activation of the enzyme sets off a chain of events within the cell that eventually leads to a response. One example of this type of enzyme-linked receptor is the tyrosine kinase receptor (Figure 9.8). A kinase is an enzyme that transfers phosphate groups from ATP to another protein. The tyrosine kinase receptor transfers phosphate groups to tyrosine molecules (tyrosine residues). First, signaling molecules bind to the extracellular domain of two nearby tyrosine kinase receptors. The two neighboring receptors then bond together, or dimerize. Phosphates are then added to tyrosine

residues on the intracellular domain of the receptors (phosphorylation). The phosphorylated residues can then transmit the signal to the next messenger within the cytoplasm.
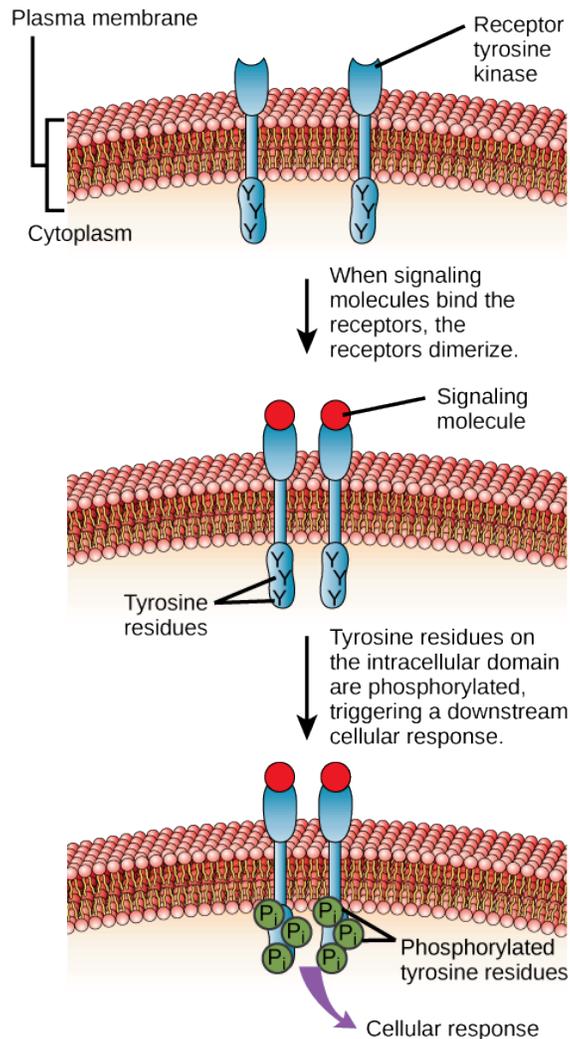

Figure 9.8. A receptor tyrosine kinase is an enzyme-linked receptor with a single transmembrane region, and extracellular and intracellular domains. Binding of a signaling molecule to the extracellular domain causes the receptor to dimerize. Tyrosine residues on the intracellular domain are then autophosphorylated, triggering a downstream cellular response. The signal is terminated by a phosphatase that removes the phosphates from the phosphotyrosine residues.

Signaling molecules

Produced by signaling cells and the subsequent binding to receptors in target cells, ligands act as chemical signals that travel to the target cells to coordinate responses. The types of molecules that serve as ligands are incredibly varied and range from small proteins to small ions like calcium (Ca2+).

Small hydrophobic ligands

Small hydrophobic ligands can directly diffuse through the plasma membrane and interact with internal receptors. Important members of this class of ligands are the steroid hormones. Steroids are lipids that have a hydrocarbon skeleton with four fused rings; different steroids have different functional groups attached to the carbon skeleton. Steroid hormones include the female sex hormone, estradiol, which is a type of estrogen; the male sex hormone, testosterone; and cholesterol, which is an important structural component of biological membranes and a precursor of steriod hormones ([Figure 9.9](#)). Other hydrophobic hormones include thyroid hormones and vitamin D. In order to be soluble in blood, hydrophobic ligands must bind to carrier proteins while they are being transported through the bloodstream.
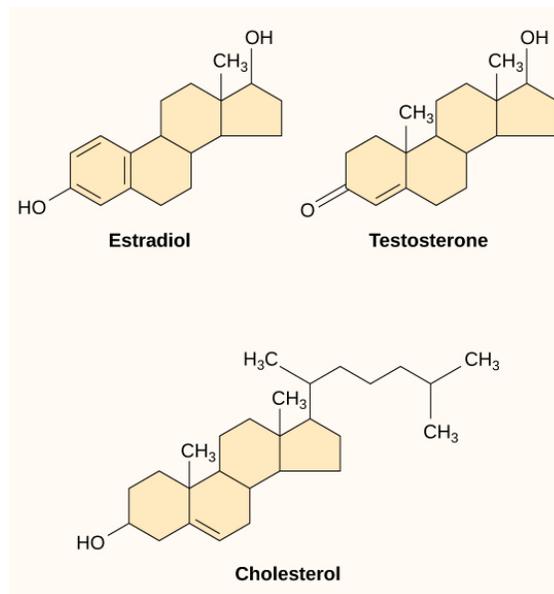

Figure 9.9. Steroid hormones have similar chemical structures to their precursor, cholesterol. Because these molecules are small and hydrophobic, they can diffuse directly across the plasma membrane into the cell, where they interact with internal receptors.

Water-soluble ligands

Water-soluble ligands are polar and therefore cannot pass through the plasma membrane unaided; sometimes, they are too large to pass through the membrane at all. Instead, most water-soluble ligands bind to the extracellular domain of cell-surface receptors. This group of ligands is quite diverse and includes small molecules, peptides, and proteins.

Other ligands

Nitric oxide (NO) is a gas that also acts as a ligand. It is able to diffuse directly across the plasma membrane, and one of its roles is to interact with receptors in smooth muscle and induce relaxation of the tissue. NO has a very short half-life and therefore only functions over short distances. Nitroglycerin, a treatment for heart disease, acts by triggering the release of NO, which causes blood vessels to dilate (expand), thus restoring blood flow to the heart. NO has become better known recently because the pathway that it affects is targeted by prescription medications for erectile dysfunction, such as Viagra (erection involves dilated blood vessels).

Hormones

Hormones mediate changes in target cells by binding to specific **hormone receptors**. In this way, even though hormones circulate throughout the body and come into contact with many different cell types, they only affect cells that possess the necessary receptors. Receptors for a specific hormone may be found on many different cells or may be limited to a small number of specialized cells. For example, thyroid hormones act on many different tissue types, stimulating metabolic activity throughout the body. Cells can have many receptors for the same hormone but often also possess receptors for different types of hormones. The number of receptors that respond to a hormone determines the cell's sensitivity to that hormone, and the resulting cellular response. Additionally, the number of receptors that respond to a hormone can change over time, resulting in increased or decreased cell sensitivity. In **up-regulation**, the number of receptors increases in response to rising hormone levels, making the cell more sensitive to the hormone and allowing for more cellular activity. When the number of receptors decreases in response to rising hormone levels, called **down-regulation**, cellular activity is reduced.

Receptor binding alters cellular activity and results in an increase or decrease in normal body processes. Depending on the location of the protein receptor on the target cell and the chemical structure of the hormone, hormones can mediate changes directly by binding to **intracellular**

hormone receptors and modulating gene transcription, or indirectly by binding to cell surface receptors and stimulating signaling pathways.

Intracellular hormone receptors

Lipid-derived (soluble) hormones such as steroid hormones diffuse across the membranes of the endocrine cell. Once outside the cell, they bind to transport proteins that keep them soluble in the bloodstream. At the target cell, the hormones are released from the carrier protein and diffuse across the lipid bilayer of the plasma membrane of cells. The steroid hormones pass through the plasma membrane of a target cell and adhere to intracellular receptors residing in the cytoplasm or in the nucleus. The cell signaling pathways induced by the steroid hormones regulate specific genes on the cell's DNA. The hormones and receptor complex act as transcription regulators by increasing or decreasing the synthesis of mRNA molecules of specific genes. This, in turn, determines the amount of corresponding protein that is synthesized by altering gene expression. This protein can be used either to change the structure of the cell or to produce enzymes that catalyze chemical reactions. In this way, the steroid hormone regulates specific cell processes as illustrated in Figure 9.1.
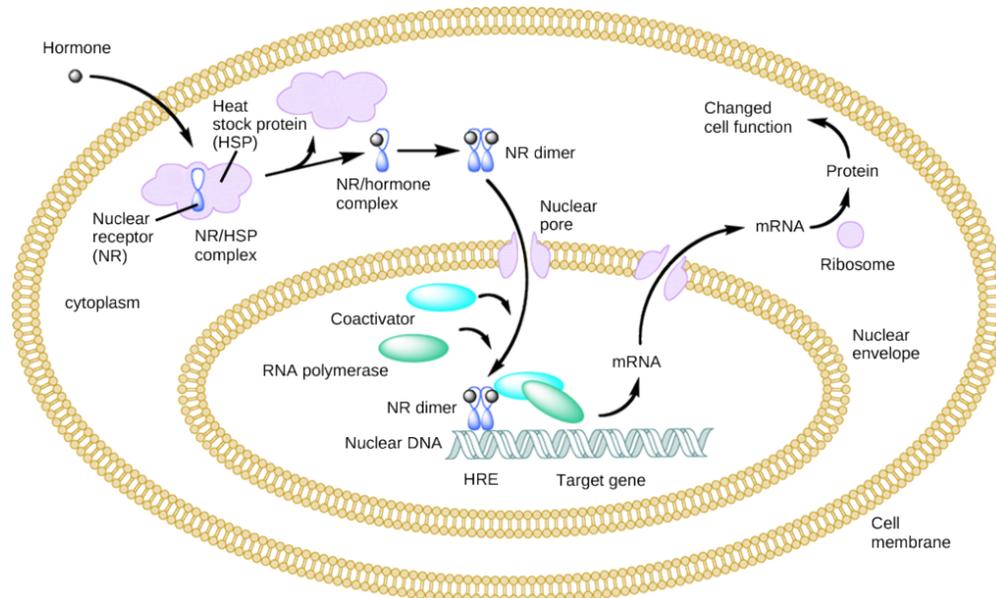

Figure 9.10. Heat shock proteins (HSP) are so named because they help refold misfolded proteins. In response to increased temperature (a "heat shock"), heat shock proteins are activated by release from the NR/HSP complex. At the same time, transcription of HSP genes is activated. Why do you think the cell responds to a heat shock by increasing the activity of proteins that help refold misfolded proteins?

Other lipid-soluble hormones that are not steroid hormones, such as vitamin D and thyroxine, have

receptors located in the nucleus. The hormones diffuse across both the plasma membrane and the nuclear envelope, then bind to receptors in the nucleus. The hormone-receptor complex stimulates transcription of specific genes.

Plasma membrane hormone receptors

Amino acid derived hormones and polypeptide hormones are not lipid-derived (lipid-soluble) and therefore cannot diffuse through the plasma membrane of cells. Lipid insoluble hormones bind to receptors on the outer surface of the plasma membrane, via **plasma membrane hormone receptors**. Unlike steroid hormones, lipid insoluble hormones do not directly affect the target cell because they cannot enter the cell and act directly on DNA. Binding of these hormones to a cell surface receptor results in activation of a signaling pathway; this triggers intracellular activity and carries out the specific effects associated with the hormone. In this way, nothing passes through the cell membrane; the hormone that binds at the surface remains at the surface of the cell while the intracellular product remains inside the cell. The hormone that initiates the signaling pathway is called a **first messenger**, which activates a second messenger in the cytoplasm, as illustrated in Figure 9.6.

One very important second messenger is cyclic AMP (cAMP). When a hormone binds to its membrane receptor, a **G-protein** that is associated with the receptor is activated; G-proteins are proteins separate from receptors that are found in the cell membrane. When a hormone is not bound to the receptor, the G-protein is inactive and is bound to guanosine diphosphate, or GDP. When a hormone binds to the receptor, the G-protein is activated by binding guanosine triphosphate, or GTP, in place of GDP. After binding, GTP is hydrolysed by the G-protein into GDP and becomes inactive.The activated G-protein in turn activates a membrane-bound enzyme called **adenylyl cyclase**. Adenylyl cyclase catalyzes the conversion of ATP to cAMP. cAMP, in turn, activates a group of proteins called protein kinases, which transfer a phosphate group from ATP to a substrate molecule in a process called phosphorylation. The phosphorylation of a substrate molecule changes its structural orientation, thereby activating it. These activated molecules can then mediate changes in cellular processes.The effect of a hormone is amplified as the signaling pathway progresses. The binding of a hormone at a single receptor causes the activation of many G-proteins, which activates adenylyl cyclase. Each molecule of adenylyl cyclase then triggers the formation of many molecules of cAMP. Further amplification occurs as protein kinases, once activated by cAMP, can catalyze many reactions. In this way, a small amount of hormone can trigger the formation of a large amount of cellular product. To stop hormone activity, cAMP is deactivated by the cytoplasmic enzyme **phosphodiesterase**, or PDE. PDE is always present in the cell and breaks down cAMP to control hormone activity, preventing overproduction of cellular products.The specific response of a cell to a lipid insoluble hormone depends on the type of receptors that are present on the cell membrane and the substrate molecules present in the cell cytoplasm. Cellular responses to hormone binding of a receptor include altering membrane permeability and metabolic pathways, stimulating synthesis of proteins and enzymes, and activating hormone release.

Question 9.2

A new antagonist molecule has been discovered that binds to and blocks plasma membrane receptors. What effect will this antagonist have on testosterone, a steroid hormone?
a. It will block testosterone from binding to its receptor.
b. It will block testosterone from activating cAMP signaling.
c. It will increase testosterone-mediated signaling.
d. It will not affect testosterone-mediated signaling.

Question 9.3

What effect will a cAMP inhibitor have on a peptide hormone-mediated signaling pathway?
a. It will prevent the hormone from binding its receptor.
b. It will prevent activation of a G-protein.
c. It will prevent activation of adenylate cyclase.
d. It will prevent the activation of protein kinases.

Question 9.4

Name two important functions of hormone receptors.

Question 9.5

How can hormones mediate changes?

Question 9.6

List all the structures that are involved in general hormonal signaling. Once you have all the structures listed, arrange them from small to large. Next, arrange them in order of events during hormonal signaling, from first to last.

9.3 – Examples of Internal Balance

9.3. Use different hormone examples (molting, digestion and circadian rhythm-related) to explain their role in maintaining homeostasis.

Maintaining homeostasis within the body requires the coordination of many different systems and organs. Communication between neighboring cells, and between cells and tissues in distant parts of the body, occurs through the release of chemicals called hormones. Hormones are released into body fluids (usually blood) that carry these chemicals to their target cells. At the target cells, which are cells that have a receptor for a signal or ligand from a signal cell, the hormones elicit a response. The cells, tissues, and organs that secrete hormones make up the endocrine system. Examples of glands of the endocrine system include the adrenal glands, which produce hormones such as epinephrine and norepinephrine that regulate responses to stress, and the thyroid gland, which produces thyroid hormones that regulate metabolic rates.

Here is an article that describes how the hypothalamus, pituitary, and adrenal gland play a role in the timing of migration and reproductive cycle of a songbird, which should help you make connections between what you have learned about hormones and endocrine system function in Section 9.2 and animal life cycle.

Question 9.8.

Find a paper that documents the effect of a hormone on an animal reproductive system. Which paper did you choose?

Question 9.9

Pick a figure from this paper that explains this hormone effect the best. Which figure did you choose and why?

Question 9.10

Show this figure to a peer in your class and ask them the following question: based on this figure what are the main results from this study? How was the data collected? What was the hypothesis that was tested / study question that was asked based on the figure? You should accept the same challenge from your peer and the figure from the paper they have selected. This type of activity is called Reverse Journal Club and it will help you learn more about the endocrine system and reading scientific articles.