Principles of Biology

Collection edited by: Robert Bear, David Rintoul, Bruce Snyder, Martha Smith-Caldas, Christopher Herren, and Eva Horne
Content authors: David Rintoul, Robert Bear, OpenStax, Steve Altaner, and Eva Horne
Online: <http://legacy.cnx.org/content/col11569/1.25>

Collection structure revised: 2016/08/09
PDF Generated: 2017/04/06 13:12:13

Table of Contents

1 | STUDIO BIOLOGY - WHAT IS IT?

The Studio Format

Introduction

> "The first principle is that you must not fool yourself - and you are the easiest person to fool."

Richard Feynman, American physicist and Nobel Prize winner, delivering the Caltech commencement address, 1974

Welcome to Principles of Biology, Kansas State University's innovative introductory biology course. Because this course is almost certainly unlike any course you have taken before, we need to spend a little time to introduce it, and tell you why this course is a great way to learn about biology.

Unlike the traditional lecture & lab introductory biology courses at most universities, Biology 198 at K-State is a studio-format course, combining lecture and lab into the same class period. There are some unique things about studio courses, and especially this one. Our studio model involves 2 separate 2-hour sessions per week, with a maximum of 78 students in the studio; thus you will spend about 4 hours per week in the studio classroom. So it is important to understand that you are in a studio course, which is not a lecture, and not a lab, but is actually a hybrid of lecture and lab. Although it is an introductory course, it was developed with input from all the faculty members in the Division of Biology. There are usually two faculty members, two GTAs and one or more undergraduate practicum student instructors per 80 students in each section.

Why do we teach this course this way? Because we believe in education, and also in giving KSU students a lot of education for their tuition dollars. The studio format has been shown to be a very effective way for us to help you learn about biology. In fact, it is about **twice as effective (http://www.lifescied.org/content/7/2/234.full.pdf+html)** as the traditional lecture/lab course in terms of your learning and retention of the material. So that's why we teach it this way.

It is also unique in that the faculty members teaching this course can include anyone in the department, including full professors. Introductory science courses, in particular, tend to be taught by graduate teaching assistants here, and at other institutions. If they are taught by a full professor in Biology 198, many freshman students will not have another course taught by a full professor until their junior or senior year. At many of our peer institutions, introductory biology courses are taught with a single instructor lecturing to 500-800 students, accompanied by a lab taught solely with graduate students. That's a relatively inexpensive way to teach introductory science courses, but also a relatively ineffective way. If you take advantage of the significant resources (both personnel and material) that the Division of Biology devotes to this course, you will learn a lot of biology. Equally importantly, you will learn how to study and be successful in a university environment. That's another advantage of the studio format!

Course materials

Two items are essential to your successful learning in this course, both of which are designed to maximize learning in the studio environment. The first is the free electronic textbook, which you are reading now. The second is the Principles of Biology Studio Manual, which must be purchased from the KSU Biology Graduate Student Association. It may look like a lab notebook, but is actually something quite different. The studio manual is analogous to your lecture notes in a standard lecture class; it is simply YOUR record of what you do in the studio. What you see, do and hear during your time in class will be recorded in your Principles of Biology Studio Manual. More importantly, it is not analogous to a lab notebook in a lab class. You do not need to turn it in to be graded (just like nobody grades your lecture notes in a lecture class!). So please treat that studio manual, which is a required text for this course, like you would your lecture notes in any other class. Read it over before the next class, mark down any questions you might have, and make sure you get a copy of the notes from another student if you have to miss a studio class.

Course and testing structure

The course is divided into 7 units, each with four or five class periods that are devoted to those units. There are tests on all 7

units, and the dates for those tests can be found on your course syllabus (link). You will start with an Introduction to Science and Biology (including two classes on Evolution), then go immediately into the study of Ecology and Ecosystems. After gaining this large-scale perspective, you will move to the study of living things at the smallest scale (molecules and cells), and then move up to the organism level (Genetics, Energetics). The final two units are more traditional (Plant Biology, Animal Biology). The final exam, at the end of the semester, is not a comprehensive exam; it covers only the last unit (Animal Biology). But all of the learning that you have done in prior units will be very important in your understanding of the concepts covered in that final exam

The test questions are all written to evaluate your knowledge of the unit Objectives. More importantly, the objectives for each class period are provided to you at the very beginning of each section of the studio manual. So if you want to know "What do I have to know for the tests?", the simple answer would be "the Objectives for that unit". You will gain an understanding of these objectives in many ways, not only from this electronic textbook, but from the studio exercises, the web pages for this class, and from discussions with your fellow students both inside and outside the classroom. In addition we have prepared Study Guides that are also based on the Objectives. For each of the day's Objectives, we provide a detailed listing of all the places (e.g. textbook, studio exercise, web page, or some combination of those) where you can get the information needed to master that Objective. If you spend your time in the studio wisely, and study with those Study Guides regularly (not just the day before the exam!), you should have a very clear understanding of the material that we think you should be learning in this class. Each question on each exam is written with one of those Objectives in mind, so it should be obvious that the Objectives are the key item on which to focus your efforts.

Your responsibilities

There will be readings in the textbook for every class day (except for the very first day of class); those reading assignments are listed in the Studio Manual, right after the Objectives for every class period. The textbook readings are an introduction to the topics for the studio exercises that day, so you need to read them BEFORE coming to class each day. In order to assure that you do the reading, there will be a short quiz over the reading material for every class period. By the end of the semester, the points for these quizzes add up to approximately the same value as a unit exam. The points should provide some incentive, but as you learn more about this course you should also figure out that learning in the studio classroom is more efficient if you have a good understanding of the material covered in the reading for that day. And more efficient learning in the classroom translates to better scores on the unit exams!

In addition, your instructors will track your attendance in the studio; attendance is required in this class. Don't worry, if miss a class if you get sick, or have to travel, your grade won't suffer. You will have three free absences (no questions asked) during the semester. But there will be point penalties, as described in the syllabus, if you miss more than three studio classes in the semester. The reasoning behind this attendance policy is quite simple. We have excellent data which prove that missing multiple classes is highly correlated with lower exam scores, and lesser learning and retention. We really want you to learn biology, but you can't do that if you are not in class. So to make sure that you take full advantage of the learning opportunities in the studio, we strictly enforce this attendance policy.

Your success in this course, both in terms of amount learned and in terms of a good grade, is assured if you understand the format of the course, do the assigned readings and attend the studio sessions faithfully, and spend at least as much time studying outside the classroom as you spend inside the classroom. You are responsible for learning, just as you are responsible in every class you take. But the difference in this class is that we do everything we can to provide the resources and the environment where learning is maximized.

Instructor responsibilities

Your studio instructors will deliver a brief (10-15 minute) introductory lecture at the start of each class day, and an equally brief wrap-up session at the end. In between those two lectures, you will be working with your fellow students on the studio exercises for that day. During that time the instructors will circulate in the studio, asking questions, answering questions, and generally helping you learn the material. Please take advantage of this incredible opportunity to interact, one on one, with your faculty and GTA instructors. If you have a question, don't be shy. If you want to know if your class notes (i.e., the stuff you are writing in the Studio Manual) are accurate, ask an instructor. Their job is to help you learn the material, and they can help a lot more if they know what your questions and concerns might be.

Your studio instructors will also be responsible for grading the daily quizzes and recording those grades, usually in a course that they set up on K-State Online. If you are not familiar with K-State Online, don't worry. It is our course management system, and it is very easy to navigate. Your studio instructors will not be responsible for writing the biweekly unit exams. Since there are 10 sections of this course each semester, it is better (and more fair) if all students in all sections take the same exams. So those exams are written by the course coordinator, and all exam questions are vetted by faculty members who are teaching in one or more sections in a semester. The first 6 unit exams are administered on Monday evenings (see syllabus for the exact dates), and the final unit exam is administered on Thursday morning during finals week. Grades for

these exams will be recorded on K-State Online as well, so that you have ready access to your grade information.

Helpful hints

- Read the assigned material before coming to class.
- Don’t skip class.
- Take advantage of the learning opportunities afforded in the studio (your fellow students, the instructors, the practicum students, the studio exercises, etc.) every time you are in class.
- Take good notes and read over your notes in the studio manual within 24 hours after every class.
- Use the study guides and concentrate on the Objectives when you study for the unit exams.
- Study a little bit every day rather than cramming the day before the exams.
- Ask lots of questions, and be prepared to answer lots of questions.
- Don’t fall behind, but if you do, make every effort to catch up as soon as possible.

2 | SCIENCE AS A WAY OF KNOWING

Science as a way of knowing

> "We absolutely must leave room for doubt or there is no progress and no learning. There is no learning without having to pose a question. And a question requires doubt. People search for certainty. But there is no certainty."

– physicist Richard Feynman, in a lecture at the Galileo Symposium, 1964.

Introduction

What is "Science"? Everyone probably has some idea of what the word means, but have you ever really thought about it? If so, here are some questions to consider.

- Is science a body of knowledge?
- Is it the same thing as "truth"?
- Is it a way to understand everything, or just a few things?
- Is it a process, and if so, can everyone do it? Or do you have to be highly intelligent, highly trained, or both, if you want to understand science?

Hopefully by the end of this course, or even by the end of this first module, you will have some good answers to those questions, and will be well on your way to thinking like a scientist (at least for this class!). Let's start with the title of this chapter – Science as a Way of Knowing. That description is from the title of a great little book by biologist John A. Moore, and is actually a pretty good answer to the question of "What is Science?" Science is both a body of knowledge, and an evidence-based process for generating that knowledge. The word itself comes from a Latin term, *scientia*, which means knowledge. But science is also about a particular kind of knowledge - knowledge about the natural world. In addition, the process of "doing science" can only help us gain additional understanding about the natural world. It is of no use to us if we want to understand the supernatural. For that we need other ways of knowing.

There are also some other aspects of science which you need to know, as you move toward a better understanding of both the scientific knowledge base and the scientific process.

Science

- requires interaction with the natural world in terms of observation, detection, or measurement.
- is objective, or evidence-based; that evidence, or a repeated demonstration of the evidence, must be available to everyone. Scientists generally don't just "take your word for it."
- requires independent evaluation and replication by others.
- leads to conclusions that are always *provisional*, i.e., they will be rejected or modified if new observations or measurements show that they are false.

There are, of course, other "ways of knowing". How do we know what we know? People who study knowledge (yes, there are such people, and they are in the branch of philosophy known as epistemology) often classify that knowledge based on the source of the knowledge. In mathematics and logic, for example, we can point to things that we know are "rationally true". In science, we focus on things that are "empirically true", i.e., based on evidence that we can see, hear, touch, etc. In religion, and, to a lesser extent, in history, we focus on "revelational truth", or knowledge that comes from another source that we accept as true, based on our assessment of the reliability of the source. So the subjects that you might study at this university can depend on different sources for the knowledge that you will be gaining. In this class we will focus, as noted above, on objective evidence obtained from observations of the natural world, and we will use some very specific terms to

describe how those empirical observations form the basis for scientific knowledge and understanding.

The process of science

So how does this process work? The processes that generate scientific knowledge are known as the **scientific method**. But even as you learn this method, it is important to realize that this is not a set recipe or process that MUST be followed in all cases. The scientific method is best understood as a statement of the core logical principles underlying how science works. The process of science always uses these core logical principles, but any individual scientific enterprise might not adhere exactly to the method outlined below (**Figure 2.1**).

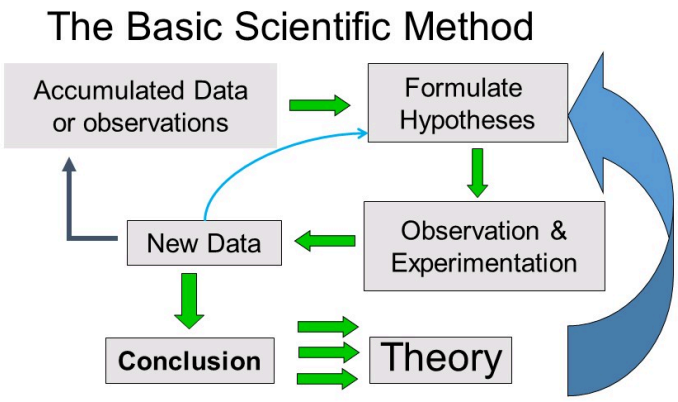


Figure 2.1 ***The Scientific Method*** – 1) Observations are used to formulate the 2) hypothesis, which is then 3) tested with experiments or new observations. The 4) new data contribute to the pool of observations, and also are used to refine the hypothesis as needed. Eventually the accumulated data allow one to make a 5) conclusion, which can contribute support for an existing 6) theory, or in some cases, support for a new theory. In all cases theories can be used to generate new testable hypotheses, which is why we say **theories are both explanatory and predictive**.

All science starts with an **observation**, or set of observations, about the natural world. You might observe a pair of male elk fighting in a high-country meadow in Colorado, for example. The next step, if you want to think about this scientifically, is to formulate some **hypothesis** to explain that observation. A hypothesis is a statement that is simply an educated guess about the cause(s) of the observed phenomenon. In order for that hypothesis to be useful in a scientific sense, however, it must have some additional characters. A scientific hypothesis must be **testable**, and it must be **falsifiable**. It does no good to generate a hypothesis that you cannot test in the real world. Thus it would not be a scientific hypothesis if you stated that the elk were fighting because invisible men in an invisible spaceship parked on the far side of the moon were controlling these elk with undetectable brain waves. That might be the actual explanation, but you can't test it, and you can't falsify it.

A good scientific hypothesis lends itself to making testable **predictions**; if the hypothesis is true, then X must be true. In this case you might state generate this hypothesis – these are male elk, and they are fighting for control of a herd of female elk. Immediately some predictions come to mind. If this hypothesis is true, you should be able to detect that these are male elk. Without getting too close, you can see that they have antlers, and previous work by other scientists (part of the set of accumulated observations that you are relying on) has shown that only male elk have antlers. Prediction confirmed. Another prediction might be that there should be one or more female elk nearby, and that these females would eventually go with the male who wins the fight or drives off the other male. You look around, and you see a herd of 10 or so female (antlerless) elk watching the spectacle. Another prediction confirmed. You will have to wait until the fight is over before you know if the prediction about the females staying with the winner is confirmed. But you have two predictions confirmed, and so far your hypothesis is supported by the evidence. More importantly, it has not been **falsified**. All of the data so far support it.

This brings up another important aspect of the process of science. In this case you made predictions and confirmed them with additional observations. You didn't do anything to the subjects; you merely observed them more closely. That is a valid approach. An equally valid approach would be to test your hypothesis by means of **experiments**. Experiments are manipulations of the experimental system, followed by additional observations. In this case, for example, you might hypothesize that the male hormone testosterone is causing the elk to fight. One prediction from that hypothesis would be that

injection of testosterone into female elk (which don't normally produce testosterone) would lead to aggressive behaviors in the female elk. You would also predict that male elk, deprived of testosterone, would be less aggressive. You might be able to come up with some other predictions from this hypothesis that could be tested with other experiments. You would have to capture some elk (male and female) to do the experiments needed to test these predictions, of course. That could be tricky, or dangerous, and you might need to hire and train help. Or you could look for similar behaviors in smaller, more easily handled animals such as mice or rats, and do the same experiments with those creatures. Both of these approaches, using observations or using experiments, are scientifically valid, as long as your hypothesis is testable and falsifiable. Furthermore, as you will learn many times in this course, there are other aspects of the experimental approach, involving concepts like **sample size, variables,** and **controls**, which you will need to consider as well. We'll save the discussion of those until a bit later, after we conclude our consideration of the general scientific method.

If you look at (**Figure 2.1**), you will see that multiple tests of the predictions lead to an increase in the number of observations. Any test of the hypothesis, no matter if it confirms or disconfirms the hypothesis, adds to our knowledge base. Generating new knowledge is one of the exciting parts of doing science, in fact. All of these observations can be used by future generations of scientists to test future hypotheses.

You'll also find another important word in (**Figure 2.1**), and that is the word **theory**. In regular conversations, people outside of science often use this word to mean an unproven or unsupported explanation, a wild guess. As you learned above, in science that description would be more appropriate for the word hypothesis. In science, a theory means something quite different. Theory is used to describe a hypothesis, or set of hypotheses, that is supported by substantial amounts of data from diverse lines of investigation. In other words, it is NEVER a "wild guess". There are many theories in science; examples relevant to the study of biology include the germ theory, the cell theory, plate-tectonic theory, and of course the grand unifying theory of evolution. All of them are well supported by incredible numbers of observations; all of them are considered the best available explanation for a diverse set of observations.

In addition, as shown in (**Figure 2.1**), you can see that just as observations lead to **predictive hypotheses**, theories can lead to predictive hypotheses as well. In fact, one of the hallmarks of a theory is that it provides a solid framework for generating hypotheses and making predictions. Scientists are confident in the explanatory power of theories, and thus are comfortable in using them to construct hypotheses, design experiments, and frame the interpretation of the data generated by those experiments. Just as a scientific hypothesis is useless if it cannot generate predictive hypotheses, a theory must serve as a framework for hypothesis building and testing. And just as the predictions of the hypothesis must be borne out by new observations if the hypothesis is going to be accepted, predictions from a theory must be supported by the observations if the theory is to continue to serve as the best available explanation for a vast set of observations.

Not shown in that figure, but implied nonetheless, is the fact that the observations must be repeatable. Other scientists, working in other locations, need to be able to do similar experiments and get the same results. That is what is meant by the statement that science is objective, not subjective. Another scientist has to be able to get the same results as you, and vice versa. Again, the history of science has thousands of examples where a new and exciting result was announced, but eventually forgotten when other scientists could not get the same result. Recent ones include the phenomenon known as "cold fusion", or the identification of a virus that was thought to cause Chronic Fatigue Syndrome (CFS). In all of these cases the original observation was found to be flawed in some way, and subsequent work, either by the original observers or by others, revealed the flaws and debunked the explanation.

Finally, it is important to remember that all scientific conclusions are provisional. In other words, a scientific conclusion is accepted as the current best explanation, but with the understanding that future investigators could make observations that might negate or modify the conclusion. So it is likely that some of the things that you will learn in this class are wrong, or at least incomplete. We still expect you to learn them, since they are the current best explanation, but it is almost certain that something in this textbook, or in the other materials for this course, will be shown by future scientists to be erroneous or incomplete. Who knows, you might be the scientist who does the work that reveals the error. Scientists actually dream about being the person who overturns a long-established notion, since that often means that their work will be remembered, and may even appear in future biology textbooks. One example of overturning a long-established concept, and ensuring a place in future textbooks, can be found in Louis Pasteur's experiments, described below

Experiments and controls

As mentioned above, a common approach to generate new scientific knowledge is to perform experiments, where the scientist changes the situation and then observes the effects of these changes. In keeping with the scientific method, this starts with an observation, from which the scientist generates a hypothesis. The hypothesis leads to a testable prediction, followed by experiments based on that testable prediction. Let's look at one of the most famous experiments in all of biology as an example.

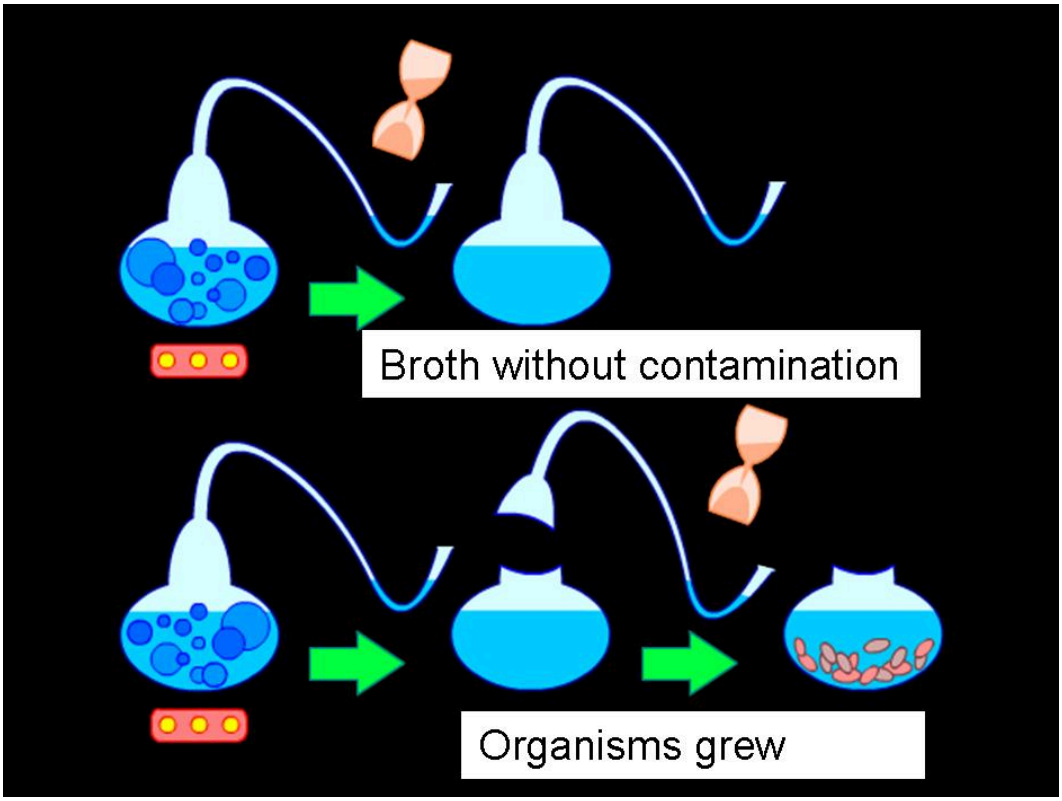


Figure 2.2 ***Pasteur's test of the hypothesis of spontaneous generation*** [By Carmel830 (Own work) [Public domain], via Wikimedia Commons]. Pasteur attempted to explain the observation that organisms (molds and bacteria) appeared in meat broth that had been boiled. His hypothesis was that these organisms came from the air, rather than from spontaneous generation. That hypothesis would predict that organisms would not appear if the meat broth was not exposed to air. He boiled the broth in flasks with long necks; air could not enter past the fluid that was left in a U-shaped section of the neck of the flask. As a control he boiled broth in other long-necked flasks, but then broke the necks off so that room air (and any microbes in that air) could fall on the broth. No organisms grew in the flasks with intact necks; organisms were found in abundance in the flasks with the broken necks.

It was widely believed in ancient times that living things arose spontaneously if conditions were right. One of the observations that led to this belief was that molds, bacteria, maggots and other life forms appeared if one left a piece of meat out in the air for a while. This concept of spontaneous generation was tested in 1860 by Louis Pasteur, using an experimental setup diagrammed above (**Figure 2.2**). Pasteur heated meat broth, in glass flasks, to a temperature where he imagined that no living things were left alive in the broth. If he left this broth out in the open, it developed active bacterial and mold growth, an observation which was consistent with the notion of spontaneous generation. But Pasteur, having recently learned about microbes, suspected that the mold and bacteria arose not from spontaneous generation, but from microbes present in the air. So he devised a set of experiments to test this hypothesis: Living microbial cells present on dust particles in the air are the source of the living cells growing in the heated meat broth.

What prediction could one make with this hypothesis? You can probably think of a couple, but the one that Pasteur came up with was that if the meat broth was in a vessel which excluded cells dropping into it from the air above, there would be no bacterial or mold growth in the broth. So he heated batches of broth in long-necked glass flasks until he thought the broth was sterilized, and also heated the necks of the flask to allow him to bend them into an “S” shape. The ends of the flasks remained open to the outside air, but dust settled in the trap in the neck of the flask and did not reach the surface of the meat broth. In other experiments, he broke off the neck after heating the flasks, allowing dust particles to settle on the broth, or waited a few days and then tilted the flasks so that the broth came into contact with the dust trapped in the bottom of the trap in the neck of the flask. Then he observed the results. Just as importantly, he repeated the experiments several times to make sure that his observations were correct.

We’ve already discussed the hypothesis, and one prediction, above. But what are some other important aspects of this experimental approach? One is the concept of a **variable**. A variable is some condition of the experimental setup that can be manipulated by the experimenter. Ideally, the experimenter should change only one variable at a time (keeping all other conditions identical); this makes interpretation of the results a lot more straightforward. What was the variable in Pasteur’s experimental setup? In this case, the variable was access of dust to the surface of the broth. In flasks that were left

open, access was allowed. In the flasks that had an intact S-shaped trap in the neck, access was not allowed. Pasteur also manipulated this variable by tilting the S-shaped flasks so that accumulated dust could contact the broth.

The other important part of this experimental approach is the concept of **control experiments**, also known by a shorter term as just **controls**. A control experiment is a setup where the variable is not introduced, so that it can be directly compared to the experimental situation where the variable (access of dust particles to the broth) is included. So a control experiment for Pasteur's incubation of broth in an open-necked flask would be incubation of broth in the S-necked flasks. If the variable is introduced by tilting the flasks, the control would again be the S-necked flasks. All other conditions (heating temperature, amount of broth, size of flasks, etc.) were the same in the experimental and control situations. The only thing that was different was a single variable (access of dust particles to the meat broth), because that was the hypothesized source of the living cells that grow in meat broth left out in the open. A single control experiment is usually all that is needed if there is only a single experimental variable being manipulated.

But it is not always possible to simplify a system so that there is only one variable. In those cases, as you will learn in the studio exercises for this class, you might need multiple control experiments. Experiments and controls will also be repeated before the investigator reports the results. It will be described in a way such that other investigators can readily repeat it as well. In some situations the results will be subjected to statistical analysis, although this was not necessary in Pasteur's experiment. Statistical analysis is critical in many scientific approaches, particularly in studies involving hypotheses about human subjects (e.g., the hypothesis that smoking causes lung cancer), where experimental manipulation of the subjects is difficult or impossible. A scientific experiment, no matter how the results are analyzed, should lead to a conclusion that either supports, or fails to support, the hypothesis. Finally, the experimental results should lead to additional hypotheses, and additional predictions, that can generate further support (or lack of support) for the hypothesis. Try to think of a few additional experiments that you might have suggested to Louis Pasteur if you were alive in 1860, and if you could speak French!

Other aspects of science

The characteristics inherent in the scientific process lead to another property of science, and that is that science is **self-correcting**. By that we mean that errors can be made, but that continued application of the tools and processes of science will usually lead to elimination of the errors and a more accurate understanding of the natural world. Science never really and finally proves anything to be true; it can, however, prove things to be untrue. To some people, in fact, that characteristic of science, its fluid and changing nature, is maddening. If you require solid ground to stand on, and immutable truth in all aspects of your life, you probably shouldn't become a scientist. If you find excitement in being part of an enterprise that is constantly changing the extent, and even the nature, of knowledge, then you have some of what it takes to be a scientist. But even if you don't become a scientist, a bit on scientific knowledge, and a bit of practice with the scientific process, will help you understand the things you need to understand in order to make intelligent decisions about many things, such as health care, climate change, or other concepts that are in your future.

Science is also a curious mix of intuitive and counter-intuitive behaviors. You practice the scientific method intuitively every day, whether you realize it or not. If you flip on the light switch in your bathroom in the morning, and the light doesn't come on, you probably take a scientific approach to solving that problem. You might hypothesize that the bulb is burnt out, and if that hypothesis is correct, replacing the bulb should solve the problem. So you do that experiment, and replace the bulb, and the light goes on, and you can continue with your daily activities blissfully unaware of the fact that you acted scientifically already that day. Intuitive science in action!

But some aspects of science, and particularly the scientific process, are not intuitive. All of us have the ability to think that our explanation of some phenomenon is correct, even if there are other observations that contradict that explanation. In fact, we often search for additional evidence to confirm our conclusion, and ignore any evidence that we might find that casts doubt on the conclusion. This is known as *confirmation bias*, and is particularly strong in situations where we have an large emotional or financial stake in the conclusion. For example, you might consider yourself a pretty good basketball player. So when you have missed 10 shots in a row, you keep shooting until you make a shot, and then you feel better about your belief that you are a good basketball player, even if those shooting percentages contradict that belief. Or you might make a visit to the chiropractor when your neck hurts, and the chiropractor might make your neck feel better. But a few days later, when it hurts again, you might not take this as a sign that chiropractic manipulation is not a cure. You might go back to that chiropractor, to have the same manipulation, because you have already invested money there, and you'd like to think that you are not wasting your money. Confirmation bias, of the active sort rather than the passive version in the previous examples, is particularly evident in people who buy into conspiracy theories. They seek out others with the same beliefs, or they only look at websites that are dedicated to that belief. It's only human nature that people like to hear what they think they already know. As a character in Terry Pratchett's Discworld series observes, "*…what people think they want is news, but what they really crave is olds…Not news but olds, telling people that what they think they already know is true*." We all like to think that we know everything that we need to know, and scientists are no exception.

So scientists have to actively guard against confirmation bias. If a scientist has an hypothesis, she has to come up with predictions and experiments that will **dis**prove that hypothesis. If the experiments indicate that the hypothesis is wrong, the scientist has to abandon that hypothesis and generate a new one, and it has to include the results of those disconfirming experiments. This is a very difficult assignment, and certainly goes against lots of human impulses. As the physicist Richard Feynman wrote, "*The first principle is that you must not foolyourself, and you are the easiest person to fool*." Every good scientist has abandoned many more hypotheses than he or she has confirmed; science teaches humility that way, for certain. Scientists must be able to change their minds if observations warrant it; there is no shame for a scientist who admits being wrong. Exactly the opposite, in fact. There are many sad examples of individuals who hung on to a hypothesis too long, and ended up with a tarnished reputation. But as we all know, admitting you are wrong is difficult for most people, and scientists are human too.

The science of biology

It's now time to shift from a discussion of science in general to the specific scientific discipline that you will be learning about, biology. Biology is the study of life. that naturally leads to the question – What is life? Suprisingly, that has proven to be a difficult question to answer in just a few words! Most textbook-level definitions of life are merely a list of characteristics; anything that exhibits all of those characteristics is said to be alive. Here's a typical list.

Living things:

- Respond to the environment.
- Assimilate and use energy from their environment.
- Maintain a relatively constant internal environment, even as the external environment changes (**homeostasis**).
- Reproduce (at the level of organisms) and can evolve (at the level of populations).
- Are highly organized, relative to their environment.

These are general characteristics, and might describe all organisms, even those which have not yet been discovered yet (e.g., those on other planets or solar systems). Until those organisms are discovered and studied, however, that statement is provisional. In addition, scientists have discovered that all living things discovered to date (i.e., the ones on this planet).are composed of one or more cells, and have DNA as their hereditary/genetic material. Some textbooks include these characteristics in their definition of life as well. More importantly, the commonality of DNA as the genetic molecule in all known life forms is strong evidence that all living things on this planet are related, i.e., they have a common ancestor. A putative common ancestor was a prediction made by Charles Darwin when he elucidated his theories about evolution and natural selection. The fact that his prediction proved to be correct is one (of many) pieces of evidence that support that theory. You'll learn about some of the other evidence later in this course.

One productive way to study and understand living things is to recognize that there is a **biological hierarchy**, which is basically an organizational concept map that allows us to focus on various levels of life.

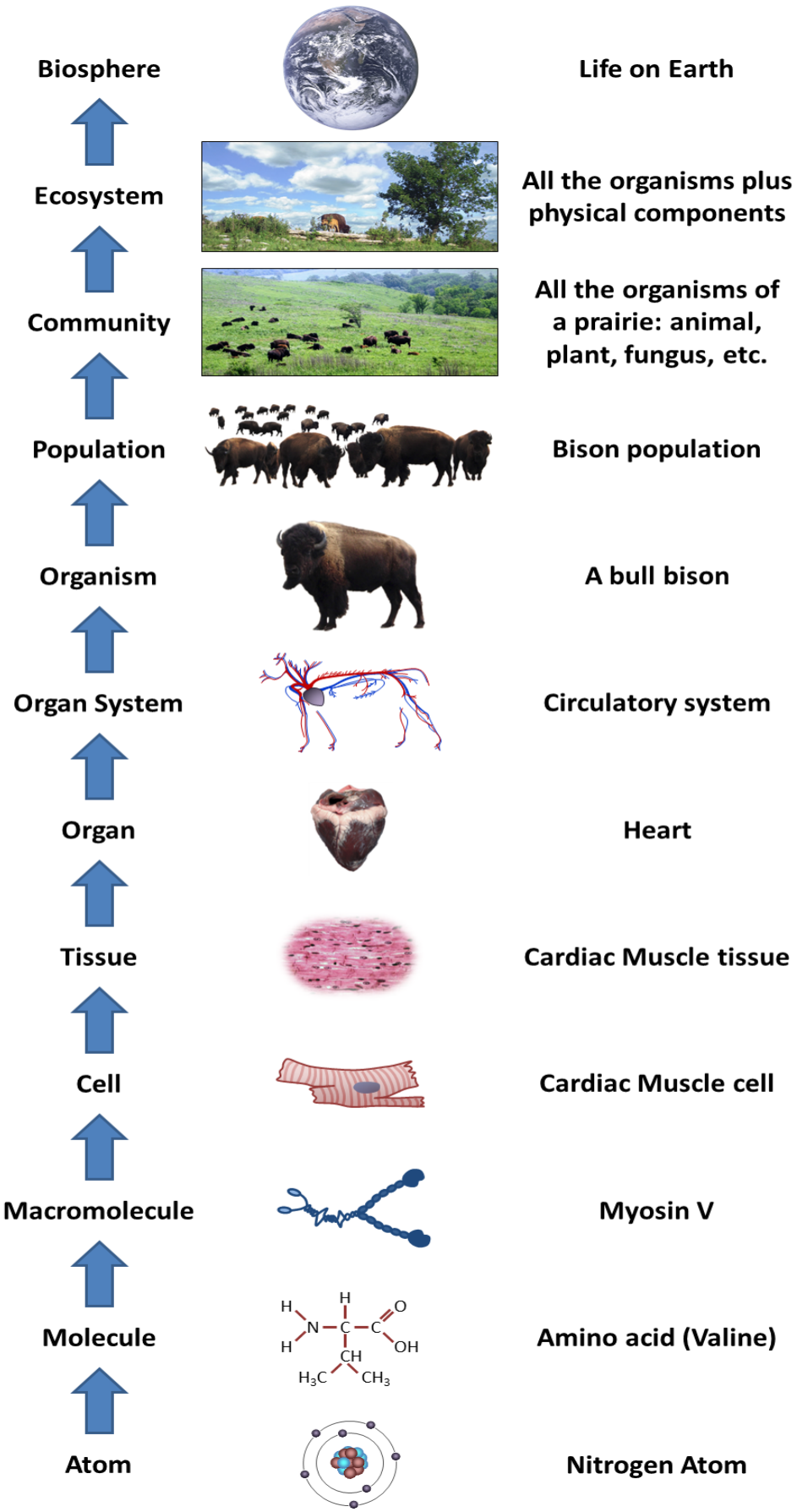

Biosphere
Life on Earth
Ecosystem
All the organisms plus physical components
Community
All the organisms of a prairie: animal, plant, fungus, etc.
Population
Bison population
Organism
A bull bison
Organ System
Circulatory system
Organ
Heart
Tissue
Cardiac Muscle tissue
Cell
Cardiac Muscle cell
Macromolecule
Myosin V
Molecule
Amino acid (Valine)
Atom
Nitrogen Atom

Figure 2.3 ***The organization of life***. Work by Eva Horne.

This hierarchy (**Figure 2.3**)extends from atoms and molecules, through cells, tissues, organs, organ systems, organisms, populations, communities, and ecosystems all the way to the biosphere (Planet Earth and the living organisms populating it). Biologists often focus on one or another of these levels, simply because it is far easier to study one level than to try to understand the entire spectrum, and the interactions between those levels. But all biologists recognize that there ARE many interactions between these levels, and those interactions lead to some very interesting and important processes as well.

Consideration of this hierarchy, coupled with the difficulty in coming up with a simple definition of life, leads some scientists to another perspective as well. These scientists argue that it is pointless to try to define life. If life arose from self-replicating chemical systems, which is the working hypothesis in the field of science known as abiogenesis, and if there is a continuum running from atoms to molecules to cells, etc., then it is not possible to point to some arbitrary place on the continuum and define it as “living”. Nobel Prize-winning abiogenesis researcher Jack Szostak writes” None of this matters, however, in terms of the fundamental scientific questions concerning the transitions leading from chemistry to biology.” Indeed, as you come to learn more about viruses in this course or elsewhere, you will probably have some sympathy for this perspective. Are viruses alive? Or would it be better to say that they are somewhere along this continuum, and bypass that question altogether?

As you proceed to learn biology in the studio classroom this semester, you will expand your understanding of the details underlying those characteristics of living things. For example, in regard to organisms responding to the environment, you will learn some of the ways that bacteria, plants and animals sense and respond to environmental conditions. You will learn how bacteria, plants and animals reproduce, and how that process of reproduction is integral to the process of evolution. You will learn about cells and tissues and organs, all of which are highly structured and organized arrangements, and how energy is obtained and expended so that these organized structures can be produced and maintained. Hopefully you will come to the realization that life, in all of its diverse incarnations, is amazing. Which is why biologists continue to study it!

3 | EVOLUTION

Evolution

> "How stupid of me not to have thought of that."

– Thomas Huxley, after reading Darwin's *Origin* (*On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*).

Introduction

What is Evolution? Surely everyone has heard the word, and perhaps a lot of other words to describe it, but do you really know what that word means, in the context of biology? Here are a few common notions about evolution. How many do you agree with?

1. Evolution has never been observed directly.
2. Evolution is only a theory, and has not been shown to be a fact.
3. Evolution means that life originated, and living things change, randomly.
4. Evolution is progress; organisms get "better" and more complicated whenever evolution occurs.
5. Evolution means that individual organisms change.
6. In order for evolution to occur, the offspring of some organisms will have to be radically different from the parental organisms.

If you said that all of these statements are false, then you have a good understanding of evolution. They are indeed all untrue. However, this is a list of some fairly common misconceptions about evolution, and many people in the world (and particularly in the USA) share one or more of these misconceptions. It is likely that you think that some or all of these statements are true. One of the hardest parts of learning is to undo a well-established misconception, so if you do think that one (or more) of those statements is true, this chapter might be a bit harder for you. But it will be worth the effort, since, as you will learn below, evolution is the guiding framework for modern biological science. Once you have a good understanding of evolution, and the mechanisms that drive it, you will be well-poised to learn and understand the biology that comes in the rest of this course.

Evolution – what is it?

The biological world is extremely diverse. In fact, that is one of the most powerful realizations that come from the study of biology, or even just from being an observant person in the world. Living things range from the microscopic bacteria to the immense blue whale. They have a diversity of life styles and metabolic capacities, from photosynthetic creatures who can make their own food from carbon dioxide gas, to predatory creatures, all the way to parasitic creatures who have some of the most complicated life styles of all. Within any one of these groups, there is also astounding diversity. Open any field guide, whether for birds, mammals, flowering plants, or mushrooms, and you will be confronted with an abundance of colors, sizes, shapes and behaviors. Even within a single species, say *Homo* sapiens, there is diversity. Look around your classroom and you will see people with a wide variety of skin colors, hair colors, eye colors, heights and weights. This diversity is a fact, and for many millennia, human beings have been trying to come up with explanations for that well-observed fact.

You are probably familiar with some of these explanations. Many ancient peoples imagined that the world was created in the form that exists now, and that blue whales, pythons, and lilac bushes were created and unchanged since that creation. This is known as *typology*; every species conforms to an ideal and unchanging type, and all members of the species are true to that type. Wolves are a type, and all wolves (within certain parameters) were considered to be similar to all other wolves, but not similar to foxes, and even less similar to lions. And all of these creatures had ancestors who were also true to the type. Once it became clear that there had been creatures, preserved in the fossil record, unlike any creatures seen today, other explanations were needed to account for these new observations. When it became clear, from geology, that the earth was very ancient, and had been in existence for millions and even billions of years, other explanations became even more satisfactory. When it became clear, from studies of comparative anatomy, that many creatures shared anatomical and developmental similarities, even though they were of different types, other explanations became obvious.

We won't go through the many explanations for the diversity of life that have been proposed and been discarded over the

centuries. There are lots of places where you can read about that historical progression, and it is interesting, for sure. Rather we will get to the explanation that is the most widely accepted scientific explanation today, and show how this explanation is supported by evidence, and also leads to predictive hypotheses that can serve as a further test of the explanation. That explanation is known as the Theory of Evolution.

So we have a lot of observations to explain, and in this case those observations are that there are a lot of diverse organisms on the planet, and that they change over time. Those are observed facts, and can be categorized under the broad umbrella of "evolution". As discussed in the previous chapter, theories are powerful frameworks for explaining observations, and for making new predictions about the natural world. The theory of evolution is no exception. It is important to understand (see myth #2 in the list above) that evolution is an observable fact, and that evolutionary theory (which is not the same thing) is the best current explanation for those observable facts. Indeed, it is the most powerful explanatory framework in biology today. Theodosius Dobzhansky, a famous biologist, expressed this sentiment quite well when he wrote in 1973, “Nothing in biology makes sense except in the light of evolution.” On a daily basis, scientists around the world are using the theory of evolution to generate hypotheses, to interpret conclusions, and to make contributions to scientific knowledge. So let’s look at that powerful explanation in more detail.

At its simplest, evolution is defined as “descent with modification”. That is joined to another concept, natural selection, to give us the first expression of the theory of evolution, published by Charles Darwin in 1859. Darwin’s genius was in recognizing, and thoroughly explaining, that descent with modification was a common phenomenon, and that selection, whether natural or artificial (e.g. animal breeding) was an explanation for life’s diversity. So let’s look more closely at natural selection, since Darwin identified it as the engine that drives the process of evolution.

Natural selection (aka adaptive evolution) is, as Darwin pointed out 150+ years ago, analogous to the process by which animal breeders produce animals with novel traits (aka artificial selection). For example, a pigeon breeder might notice that one of his pigeons has an unusually large ruff of feathers around its neck. He breeds this pigeon with another pigeon, and selects the pigeons with the biggest ruffs from among the offspring to be the parents of the next generation. After a few cycles of this, some of the pigeon offspring will have very unusual and pronounced neck ruffs, and will look nothing at all like the original pigeon ancestor in that regard (figure 1, below). This common practice gets its name from the fact that the breeder *selects*, or chooses, specific animals to be the parents of the next generation. And it works; there are many examples where substantial changes in animal appearance, or behavior, can be brought about in just a few generations by applying this artificial method of selection.

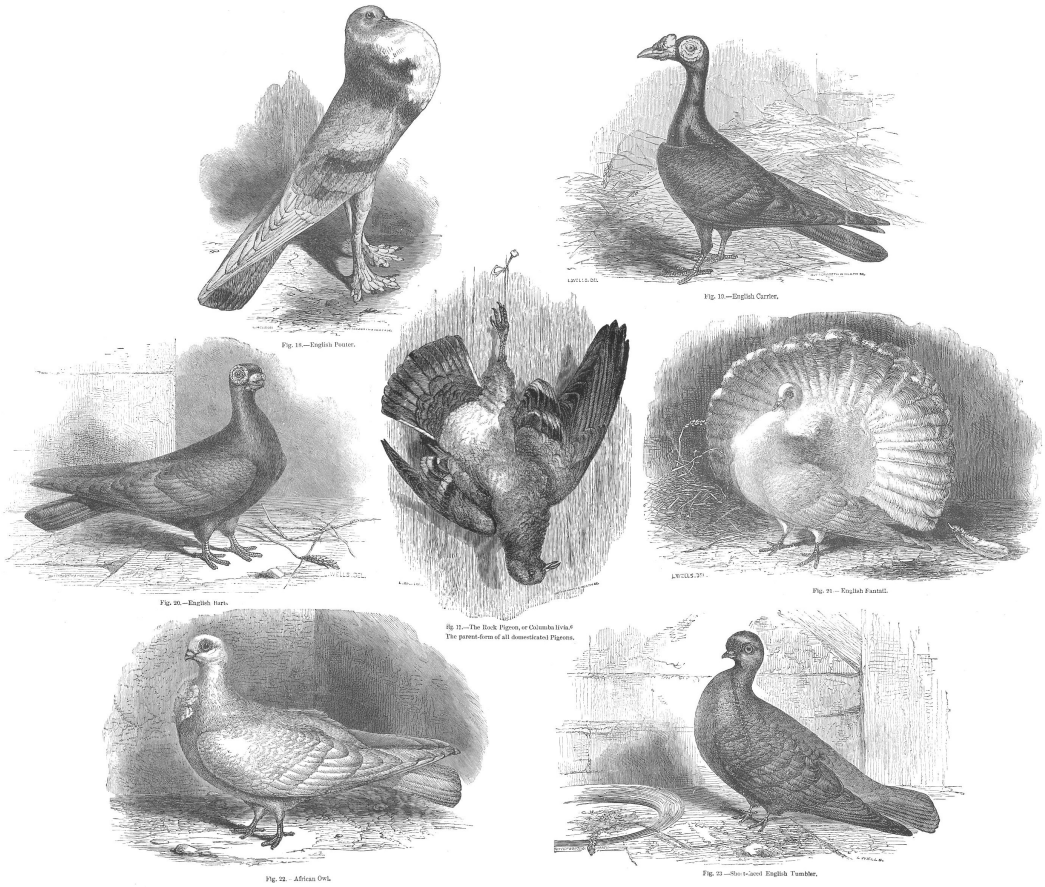


Figure 3.1 ***Darwin's pigeons*** (Original line drawings from Darwin's "Variation in Animals and Plants under Domestication", 1868). The common ancestor for all of these fancy pigeons was the Rock Pigeon (center). By selecting for unusual morphological characteristics, pigeon breeders are able to develop all of these unusual pigeons, and many more.

Darwin's genius, and the source of Huxley's self-disparaging statement at the top of this chapter, was to recognize that this process could also occur in the absence of an individual who did the selecting. *Natural selection*, the idea for which Darwin is so famous, simply recognizes three well-known observations and puts them into a context that generates evolutionary change. Let's look briefly at each of these three observations.

The first thing that Darwin postulated is that the variations seen in living things are due, to a greater or lesser degree, to heritable factors. In other words, there are *heritable variations* among the individuals in a population of organisms. Let's break down that term a bit, and look at each of the words, using examples mostly from human populations.

Firstly, we know that there are *variations* among individuals in a population. Look around your classroom, or at your family picture album. You probably don't look exactly like your brother or sister, and your mom and dad don't look exactly like your uncles or aunts. So even in situations where the parents are the same, variation occurs among the offspring. Variation is even greater in a population of individuals who don't share the same parents. Variation is normal, and easily observed.

What about that other word, *heritable*? Again we now know that many of those variable traits are heritable, i.e. they are passed from one generation to the next. In humans, eye color, hair color, height, etc. are all characteristics that might be the same in you and your parents. If you have a dog or cat, and that dog or cat has offspring, you can often see aspects of the offspring (e.g. coat color, size) that are identical to those in the parental animal. One likely explanation for that observation is that you and your parents have some shared molecule or molecules that determine each of those traits. We now know (but Darwin didn't) that the molecule is DNA, about which you will learn more later. On the other hand, some conditions are not heritable. For example, if you have a cat that lost its tail in a horrible and noisy accident involving a rocking chair and your 300-lb great-aunt, and if the cat has kittens, those kittens will have normal tails. The rocking chair might damage the cat's tail, but not its DNA. At the time of Darwin, the mechanisms of heritability were not known (he knew nothing about genes), but everyone understood that some traits were heritable, and others were not. So again, the heritable variation that is necessary for evolution to occur is easily observed in the natural world.

The second thing that Darwin observed, and that was a huge factor in his synthesis of these observations into his theory,

is that not all of the individuals in a given generation will survive and reproduce to the same degree. Simple mathematics corroborates that. If all of the fruit flies from a single pair of fruit flies survived and produced a maximal number of offspring, after a mere 25 generations (which can take just a single year in this species) that population of flies would fill a ball 96 million miles in diameter, or more than the distance from the earth to the sun. Fruit flies have been around for lots longer than a year, and you can still see the sun, so obviously fruit flies do not all survive and reproduce.

Finally, the third condition necessitates that these heritable variations can result in differences in survival or reproductive success. Again, there is abundant evidence for that. Inherited human conditions that result in mental retardation, or physical deformation, often mean that the affected individual will not survive or reproduce. Medical intervention has, in some cases, been able to counteract those disabilities and allow individuals with some inherited conditions to survive and reproduce, but in previous generations, or in populations of organisms that do not have access to medical care, many heritable variations were not represented in the next generation because the individuals with those variations simply did not reproduce.

So the model Darwin proposed is quite simple. If all of those conditions were true, organisms with heritable variations that enhanced their chances for survival and reproduction would be more likely to be among the parents of the next generation, and the frequency of those organisms with those particular heritable variations would increase in the next generation. This is a simple idea, but it has many ramifications for the study of biology.

It seems clear, just from observations we all have made, that these three conditions do pertain in the natural world. If that is the case, then the process of natural selection could operate, and variations that resulted in reproductive success would become more common in the population. It is important to understand that this process is the result of an interaction between the organisms and their environment. Over time, organisms that fit better into that environment will become more abundant in the population, and may eventually be the only organisms in the population. The term *fitness*, in this context, simply is a measure of how well individuals with certain traits survive and reproduce in a particular environment. The environment is an incredibly important aspect of this process. If the environment changes, organisms which were fit for the previous environment may suddenly find themselves less well-adapted, and rare organisms that were ill-adapted in the previous environment may suddenly become more fit to that new environment. Fitness is relative, and the environment is a major player in the determination of fitness.

In addition, consideration of these processes in the real world leads to a better understanding of the questions in the introduction to this chapter. As you can see, the process of evolution is NOT random; the interaction of the organism and its environment leads to selection, and selection, by the very nature of the word, is not random. Just as an animal breeder chooses specific individuals as the parents of the next generation, the process of natural selection chooses specific individuals as the parents of the next generation, leading to evolution of the population. There are some important differences, however. In artificial selection, the breeder has a goal (e.g. to get a goat that produces more milk), and designs the breeding program with that goal in mind. In natural selection, there is no ultimate goal, and no plan; organisms are selected for their adaptation in a particular environment, which can (and often does) change. The process is unguided, in the sense that there is no goal in mind, but unguided is not the same thing as random.

Secondly, careful consideration of this process also disproves the notion that evolution equals progress toward a “better” organism. An organism that is better adapted to one environment can be very ill-adapted if the environment changes. In that situation, a “worse” organism, one that is rare in the first environment, is now the “better” one in the new environment. That is not progress, it is just change. In fact, some organisms become so well-adapted to their environments that they lose some of the complex structures or pathways that their ancestors had. Cave fish have no eyes, even though their ancestors did. Whales have no legs, even though their ancestors did. Some parasites, living in a rich sea of nutrients, have lost organelles such as mitochondria, even though their ancestors had those organelles and all of the metabolic pathways associated with them. These highly-adapted organisms are actually less complex than the ancestors from which they evolved. Evolution clearly is not a synonym for progress!

Finally, it should be clear that evolution is a change at the level of the population, and not at the level of the organism. Natural selection acts on organisms, but the result of selection is seen in the next generation. And this change is usually very gradual; there is no need to invoke absurd situations where a cat gives birth to a dog, or vice-versa.

Darwin correctly pointed out the analogies between this process of natural selection and artificial selection, the well-known process that animal breeders used to select for interesting or useful variants in animal species. In other words, natural processes can generate the diversity we find in the natural world if all of those conditions are true, and if there is sufficient time to produce many generations. You will learn more in the studio exercises about how even small differences in reproductive success can, over time, lead to large changes in the characteristics of organisms in a population. Small changes (one or two genes in organisms that still are members of the same species) are sometimes described as *micro-evolution*. Larger changes that result in different species, for example, are described as *macro-evolution*. This is an artificial distinction, actually. Macro-evolution is merely micro-evolution that has proceeded for a longer time. For a clever graphical illustration of that, see figure 2 below.

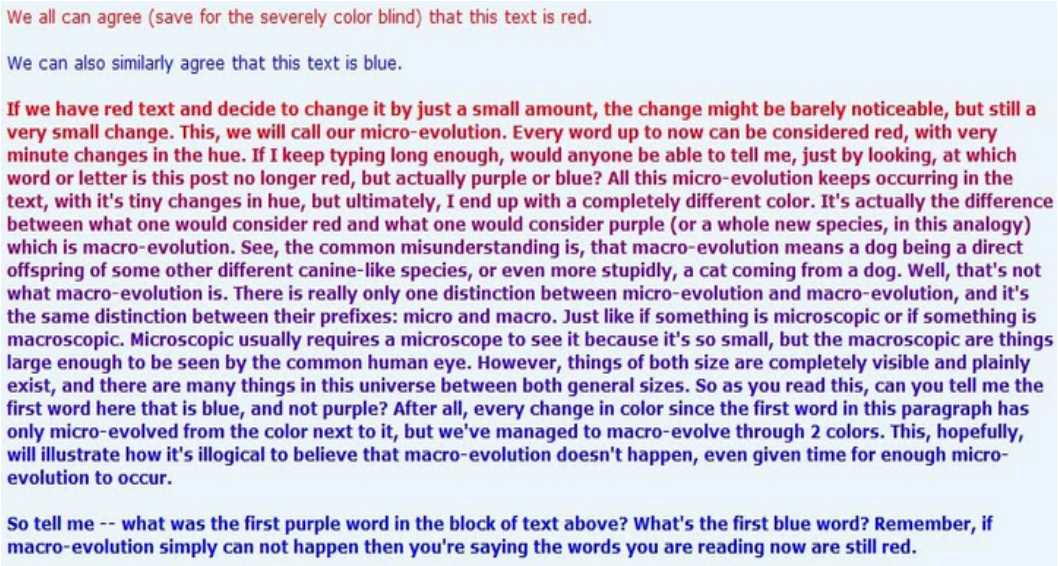


Figure 3.2 ***How very gradual changes can, over time, result in significant changes: A textual example***.

In summary, natural selection is a powerful agent, and recognition of this process was a powerful insight. Darwin proposed his theory in 1859, and elaborated on it in other books and other editions of the *Origin of Species*. Since Darwin's time, other scientists have identified other agents, in addition to natural selection, that result in changes in the characteristics of a population, and you will learn more about those in later chapters. Additionally, other scientists made many predictions based on this explanatory framework, and did many experiments to test those predictions. Scientists are still engaged in that process today, and Darwin's ideas have been confirmed many times over, and even extended so that we understand how the process works in much more detail than Darwin did. That is, as you learned in the previous chapter, one hallmark of a great theory.

Evidence for Evolution

There are multiple lines of evidence, many of which were unimagined in the time of Darwin, that support his explanation for the diversity of life. The following is not meant to be an exhaustive cataloguing of that evidence. Indeed, more evidence accumulates every day, making it impossible to point out all of the threads in that fabric. It is also important to recognize that the evidence doesn't come just from biology. For example, as noted above, Darwin's explanation would require a lot of time and many generations. If the earth was too young, none of this could have happened. The sciences of physics and geology confirm that the earth is over 4.5 billion years old, which is plenty long enough for evolution to occur. The fossil record, the research subject of geology and paleontology, also provides substantial supporting evidence for Darwin's big idea. The discovery of continental drift, and the development of plate tectonic theory, made sense of a lot of observations about both the fossil record and about populations of living organisms. Let's look at a few of the lines of evidence, and see how they all weave together to make the coherent and elegant fabric that is the hallmark of a good scientific theory.

The fossil record

In science, radically new explanations can only be successful when the conventional explanations no longer explain all the observations. In the history of biology, this was the situation in the early part of the nineteenth century, when many interesting fossils were being discovered and carefully scrutinized. It soon became apparent that fossils were indeed the remains of once-living organisms, and that fossils in geologically younger strata seemed to be both similar and different from those in older strata. The fossil record showed that whole groups of organisms appeared and disappeared during the history of the earth. Others seemed to be much the same in rocks of different ages. Familiar organisms, particularly marine mollusks such as clams and snails, could be found in older rocks, but in many cases these organisms were not identical to the current organisms. Plant fossils told the same story. The reigning explanation for the diversity of life, creation of all these creatures at the same time and place, clearly did not explain these new observations. Evolutionary theory was a much more satisfying scientific explanation, and the development of that theory by Darwin and others started at that time.

Since Darwin's time the fossil record has become much more extensive, and the evidence for this explanation has become much more well-supported. Gaps in the fossil record that were pointed out by Darwin's contemporaries have been gradually filled in. Indeed, the sciences of geology and paleontology, in combination with biology, have allowed scientists to make predictions about where, exactly, particular fossils in particular gaps should be found.

The most recent (and spectacular) example of this was the discovery of fossilized remains of a creature that bridges the gap

between fish and amphibians, which the first four-legged creatures (aka tetrapods) to move onto land. The fossil record, coupled with genetic evidence from modern-day amphibians and fish, indicated that this transition to land occurred about 375-400 million years ago. But only fish fossils, or amphibian fossils, had ever been found. Logic dictates that there should be a transitional creature, or "missing link" in popular jargon, which had characteristics of both fish and amphibians. It was reasoned that creatures such as this, if they existed, would probably live in shallow areas at the edge of seas or bays. Geologists knew which particular rock formations resulted from those sort of environments of that age, and so expeditions were dispatched to search for such fossils in one of those geological formations. These rocks were deposited in warm shallow tropical seas 375 million years ago, but are now, as a result of continental drift, located on Ellesmere Island in the Canadian Arctic. In 2004 fossils were found in those rocks that elegantly fit that prediction; the creature was named *Tiktaalik roseae*. (figure 3, below). The genus name for this "fishapod" comes from the name of a fish and was suggested by local Inuits on Ellesmere Island, and the specific epithet "roseae" honors an anonymous donor who helped fund these grueling expeditions to the high Arctic. *Tiktaalik* "fins" have basic wrist bones, but no digits, or fingers. It is truly a missing link, and its discovery stems directly from predictions made on the basis of previous scientific observations, in a classic example of the power of the explanatory framework known as the theory of evolution. Descriptions of the expeditions, and lots more about the incredible insights that have come from those fossils, can be found in a charming book called "Your Inner Fish", written by Neil Shubin and published in 2008.

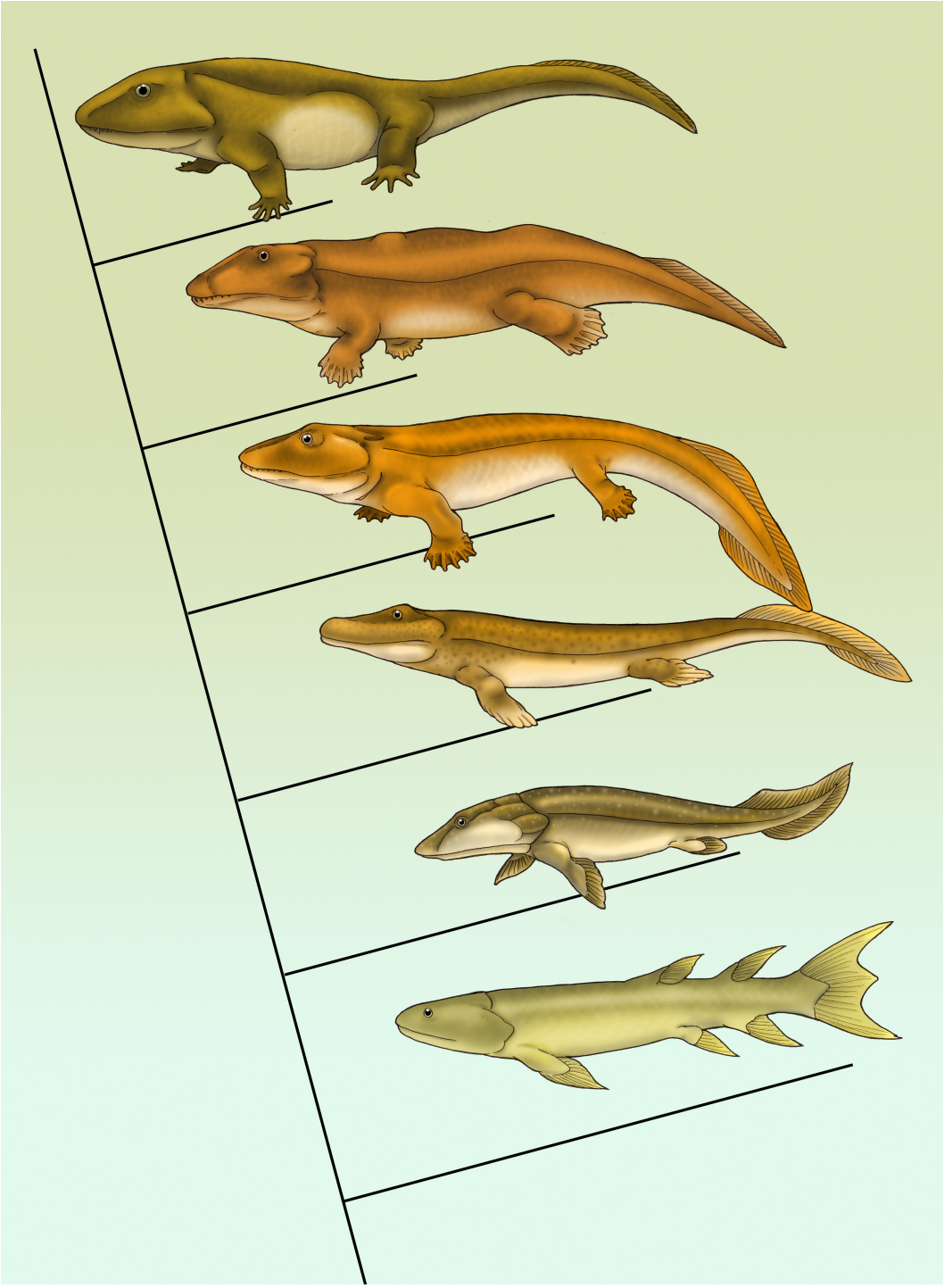


Figure 3.3 ***"Fishapod" evolution*** (By Maija Karala (Own work) [CC-BY-SA-3.0], via Wikimedia Commons. A cladogram showing the evolution of tetrapods, using the best-known transitional fossils. From bottom to top: Eusthenopteron, Panderichthys, **Tiktaalik,** Acanthostega, Ichthyostega, Pederpes.

Comparative anatomy and embryology

At the same time that the fossil record was making some scientists scratch their heads and question the creation explanation for the diversity of life, other scientists were looking more closely at these fossils and at the bones of existing organisms. These comparative anatomists also made observations which were much more easily explained by the theory of evolution. The different bones in fossil skulls, for example, could be compared to the bones in modern skulls, allowing anatomists to discern that the fossil skulls and the modern skulls had remarkable similarities in the number and the position of the individual bones in the skull. Most of the bones in a fossil fish skull have counterparts not only in modern fish skulls, but

in fossil and modern amphibian skulls, or fossil and modern reptile skulls, and even fossil and modern mammal skulls. Occasionally the fossil record shows us when a skull bone is added or one is lost, and also allows us to track progressive modifications in the positions of these bones on the skull surface. We can only understand these observations in the light of evolutionary theory – if we conclude that the bones reflect the fact that each kind of organism is descended from some other. Descent with modification is the most satisfying scientific explanation for these observations.

The anatomy of modern organisms also reflects this common ancestry. The limbs of all tetrapods contain a similar number and arrangement of bones, even though the size and shape of the bones can vary greatly in different organisms. For example, the two bones in your forearm, the radius and the ulna, have counterparts in other mammals (figure 2.4), in reptiles, in birds, and even in fossil dinosaurs and pleisiosaurs . If all of these structures were specifically created for moving around in a different environment (e.g. water for the plesiosaur and air for the bird or bat), simple engineering principles would dictate that different structures would be more efficient in those different situations. Yet the same structures, endlessly modified, are found in all of them. The simplest explanation for this is that the organisms share a common ancestor where that structure originated, and evolutionary mechanisms resulted in the modifications in size and shape that we see today. This phenomenon is known as *homology*; structures are said to be *homologous structures* if they occupy similar positions and arise from a common ancestral structure.

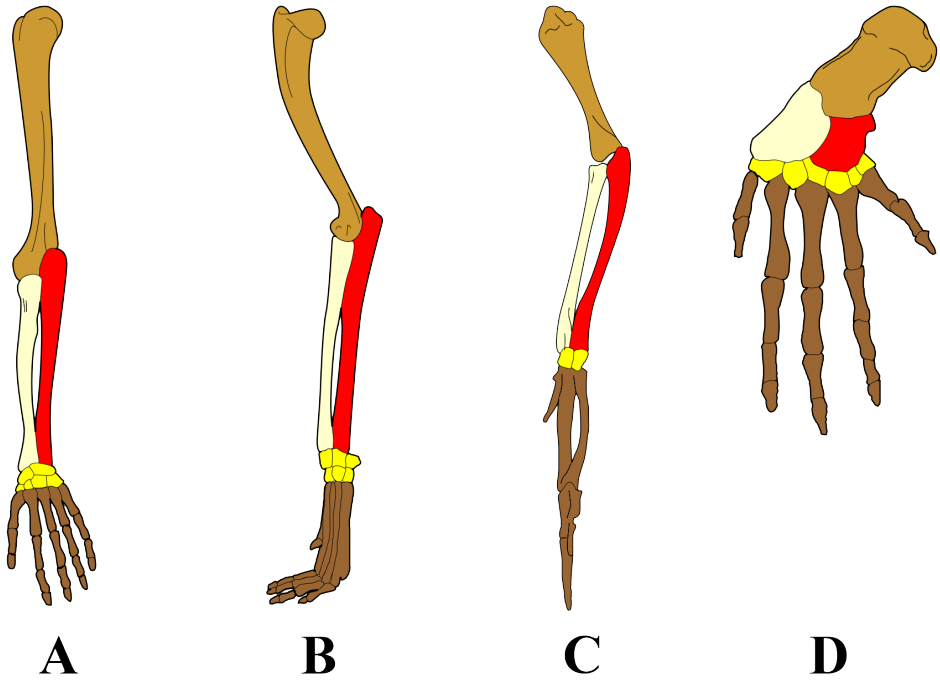


Figure 3.4 ***Homologous bones in the forelimbs of four vertebrates*** (By Petter Bøckman, via Wikimedia Commons). A-human, B-dog, C-bird, D-whale. The various colors indicate bones of various groups (e.g. dark brown = bones of the fingers,yellw = bones of the wrist, red = ulna, beige = radius, and light brown = humerus). The various bones in the forelimbs of four vertebrates differ in size and shape, resulting in very different morphologies of the forelimbs of these organisms. But both the number of bones, and their position relative to each other, are quite similar, as is their embryological development. These homologous parts provided one of Darwin's arguments in support of his theory of evolution.

Even vertebrates who have lost these limbs in the course of evolution (e.g. snakes) had similar structures prior to that evolutionary change. Figure shows the fossilized remains of a creature (*Tetrapodophis amplectus*) that lived in what is now Brazil 120 million years ago. It had a snake-like body and may be the ancestor to all snakes, but it also had four small limbs. The forelimb shown in that figure clearly has the humerus, radius, ulna and hand bones that are found in modern vertebrates.

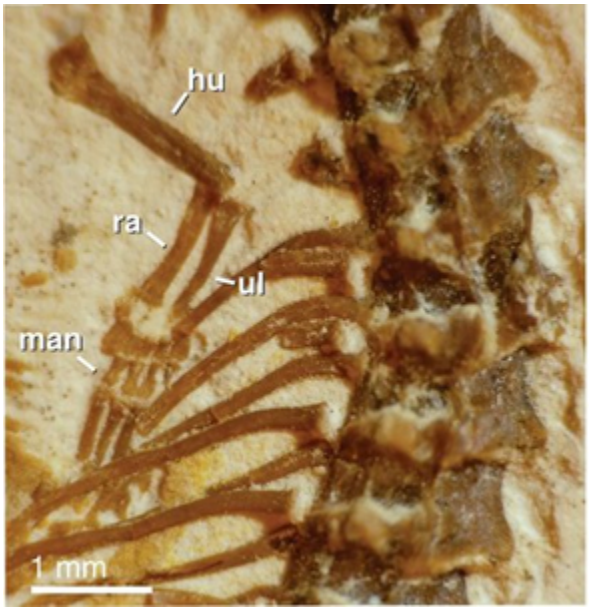


Figure 3.5 Homologous forelimb bones in a fossil snake Forelimb of *Tetrapodophis amplectus*, a four-legged snake from the early Cretaceous (120 million years ago). Hu = humerus, ra = radius, ul = ulna and man = manus (hand). Photo from Martil, D. *et al, Science* 349:416-19 (2015).

Embryologists also made predictions based on this evolutionary explanation. They predicted that homologous bones would arise from similar structures during the development of the embryo. For example, the forearm bone that we call the radius, which looks radically different in the forearms of a bat or a human or a mouse or a bird, would come from similar structures in the embryos of bats, humans, mice or birds. Those predictions also were found to be correct. So homology argues strongly for an explanation that invokes descent with modification.

In contrast, the wings of insects and the wings of bats or birds do not have similar structures, although they have similar functions (to propel the organism through the air). These structures are said to be *analogous* rather than homologous; they share a function but do not arise from a structure that is found in a common ancestor. Indeed, if organisms predominantly had analogous structures, which would be different engineering solutions to a common problem, that evidence would be more consistent with the explanation of independent creation of those organisms. But homologous structures seem to be the much more common observation, making descent with modification a much more scientifically satisfying explanation.

Comparative biochemistry

One of the biggest surprises of modern biology came from the field of science known as biochemistry. Once biochemists started to unravel the mysteries of metabolism, the unity of life on this planet became quite obvious. Creatures with incredibly different morphologies, habitats, and lifestyles all seem to have incredibly similar metabolic pathways. Bacteria, bonobos, bats and bananas all use a molecule known as ATP (adenosine triphosphate) to store and provide energy within their cells, for example. The metabolic pathway known as glycolysis, which you will learn about in subsequent chapters, is found in all the organisms on the planet, and the enzymes that are used in that pathway are quite similar in these diverse organisms. Again, this argues strongly for common ancestry, which is a strong prediction that arises from a “descent with modification” explanation. Once an ancient cell developed these metabolic pathways, there was no need to re-invent that wheel. It is somewhat ironic that some of the best evidence for a particular explanation for the diversity of life comes from the discovery of the unity of life at the molecular level.

Genetics and genomics

Besides ATP, another molecule common to all life forms on the planet is DNA (deoxyribonucleic acid). This molecule stores genetic information, so it is the molecule of heredity. Its role in heredity also means that it can be modified under some circumstances, thus giving rise to the variations described above. Darwin knew nothing about DNA when he proposed his theory in 1859; his ideas about mechanisms of heredity were, in fact, spectacularly wrong. But the discovery of the mechanisms of heredity, starting with Mendel in 1866, and extended by many others in the early part of the 20th century, made it possible to finally propose mechanisms by which heritable variations arise and are transmitted between generations. In fact, the first 4 decades of the 20th century were the years when the two seemingly unrelated fields of genetics and evolution were united. This Neo-Darwinian synthesis, starring Theodosius Dobzhansky, Ernst Mayr, and George Gaylord Simpson, resulted in modern evolutionary theory, and allowed scientists from both genetics and evolutionary backgrounds

to work together to make and test predictions.

The elucidation of the structure of DNA by Watson and Crick in 1953, followed soon by the breaking of the genetic code, provided even more evidence for descent with modification. DNA, as you will learn later, functions as a repository of information. In order for the information to be used to build a cell or an organism, it must be read and translated into different molecules. The processes, and the enzymes, that do this work of reading and translating are virtually identical in all living creatures on the planet. The genetic code was, perhaps prematurely, called the “universal” genetic code for precisely that reason; it is translated identically in almost all organisms that have been discovered to date. Once again, this is a strong argument for common ancestry and descent with modification.

But the really impressive outcome of this fusion of molecular knowledge and organismal knowledge comes from the study of the structure of genes, and genomes, at a detailed level. Incredibly, scientists have discovered molecular fossils of a sort – stretches of DNA which are not used in modern organisms, but which remain in the genome as a record of functions in the past. For example, chickens don’t have teeth, but they have genes for tooth proteins, turned off long ago, still lurking in their genomes. Those genes can be turned on under the right conditions, producing toothy structures, which were last seen in dinosaurs, the extinct ancestors of modern chickens. There is no good explanation for these observations, other than descent with modification. Similarly, detailed analysis of the DNA of organisms, including now some long-dead organisms like mammoths and Neanderthals, allows scientists to test predictions about common ancestry, and gain insights into the course of evolutionary change in all organisms. In fact, evidence from analysis of DNA, and other molecules, has allowed us to fine-tune our hypotheses about ancestry and relationships throughout the biological world, as explained in the next chapter on Taxonomy and Phylogeny.

4 | TAXONOMY AND PHYLOGENY

Taxonomy and Phylogeny

> "Birds in a way resemble fishes. For birds have their wings in the upper part of their bodies, and fishes have two fins in the front part of their bodies. Birds have feet on their under part, and most fishes have a second pair of fins in their under part…"

– Aristotle (384-322 B.C), *De Incessu Animalium*.

Introduction – Differences and Similarities

Observations and speculations about the similarities and differences of the life forms around us clearly have a long history. Aristotle's ancient musings about animals pre-date the concept of homologous and analogous structures, which we discussed in the last chapter, but his insights are accurate today. Aristotle was the first to write about his attempts to classify animals into groups, and his classification scheme was the standard for many centuries. Attempts to refine the classification of animals (and other living things) continue even today, as you will see. Those classification schemes, throughout the centuries, have used many different criteria for separating living things into different groups. "How is this thing different from this other thing?" has been the focus of many scientific endeavors. But, as Aristotle recognized in the passage above, it is just as important to ask about the similarities, and not just focus on the differences.

We use the words **taxonomy** or **systematics** to describe the activity of classifying and naming living things. There are many ways to divide living things into groups; the ability to recognize and classify things is a deeply-seated and oft-used human activity. Some of these schemes are based on habitat, e.g. water-dwelling creatures vs. land-dwelling creatures or aerial-dwelling creatures. Others are based on internal characters. For example, Aristotle's two most basic groups were those with blood and those without blood, a grouping scheme that coincidentally neatly separates most of the vertebrates from most of the invertebrates. But most schemes have been based on *morphology*, such as size, shape, number and proportion of appendages, etc. This sort of classification seems to be easy enough to do, but, as you will see later in this chapter, it can lead to some interesting mistakes.

Finally, it is important to understand that **all classification schemes should be viewed as simply being hypotheses**. Like any hypothesis, a classification scheme should change, or even be discarded, if new observations contradict the predictions of the hypothesis. This leads to some frustration on the part of some students, because (again) they would like to have some certainty about what they are learning. But a science where everything is certain would be a dead and dusty science, which certainly doesn't describe the state of taxonomy today, or tomorrow.

A brief history of taxonomy

Figure 4.1 *Taxonomic information for the Bald Eagle (photo by D. A. Rintoul)*

After Aristotle, there was not a lot of progress in taxonomy for many centuries. In fact, there may have been negative progress for some of that time, as Aristotle's system was brushed aside or forgotten. But in the 1700's a Swedish biologist who went by the Latin name Carolus Linneaus developed a system of biological classification that still underlies the system used today. His big contribution to the discipline was to introduce the concept of using two names to describe the smallest unit of classification, the *species*. In the Linnean system, every organism has a unique "scientific name", consisting of a specific epithet preceded by a name for the next highest level of classification, the genus (plural = genera). Higher levels of classification included, in order above the genus, family, order, class, phylum, and kingdom. Subsequently another top level, the domain, was added to this hierarchy, giving us the classification scheme shown above (**Figure 4.1**). This bald eagle has the scientific name *Haliaetus leucocephalus*. There are 7 other members of the genus Haliaetus. The genus is placed in the Family Accipitridae, in the Order Accipitriformes, in the class Aves, in the Phylum Chordata, in the Kingdom Animalia, and in the Domain Eukarya. Biologists refer to a group of organisms, at any level, as a **taxon** (shorthand for taxonomic unit, plural = taxa). Thus a species is a taxon, as is a Genus, or a Family, or an Order, etc.

There are an estimated 8-9 millions species on our planet. Every organism that has been described fits somewhere in this classification scheme. You can explore this in much more detail at the **Tree of Life website (http://tolweb.org/tree/)** . There you will find information on the thousands of species that have been described and named, as well as their currently assigned place in the taxonomic scheme. It is also worth pointing out that the basic unit of this scheme, the species, is not a well-defined term at all. There are multiple definitions of species that you can find with just a minimal bit of effort, and none of them is entirely satisfactory. Just as we learned about the definition of life, it is sometimes difficult to pin down an exact point on a hierarchy. As you will see from the material below, the problems with a definition of species might be related to the fact that evolution is ongoing, and that some groups of organisms are at different points on an evolutionary path at the present time.

However, if we ignore the vexing issue of figuring out a widely-accepted definition for "species", it is clear that a system based on this unit has several advantages. One of the immediate advantages of the Linnean system was that it allowed biologists to bypass the often confusing different common names that are used in various parts of the world, or in various languages. For example, the animal that is known as a moose in North America is confusingly called an elk in Europe. But when you refer to this creature by its scientific name (*Alces alces*), there is no confusion. That is why scientific papers always include the two term scientific name for a species. The hierarchical nature of the Linnean scheme is also useful, allowing us to easily understand the levels of similarity at different levels in the scheme. For example, organisms in the same genus can be very similar to each other, and may even be difficult to tell apart unless you are an expert. The similarity becomes less apparent for higher taxa; the class Aves includes that Bald Eagle mentioned above as well as your pet parakeet, which may screech like an eagle, but which is otherwise not very similar to an eagle at all.

The original Linnean taxonomy was based, as noted above, primarily on external morphology. It also relied on the obvious fact that organisms are related to each other in many ways, and that some seem more closely related than others. However, a system based solely on morphological similarities and differences can lead to some interesting classification errors. For example, Linneaus had one Phylum named Vermes (Latin for "worms"). This obsolete taxon included animals that we now classify with the mollusks, others that we now classify with the vertebrates, and still others that we now classify with crustaceans, as well as those that are still classified as worms. This reinforces two aspects of taxonomy that are important to remember. The first is that taxonomic classification schemes are merely hypotheses, and should be discarded or modified if newly obtained information is not consistent with that hypothesis. The second is that classification schemes based only on

one type of information (in this case, external morphology) can be quite mistaken. Using more than one type of information will lead to better and more stable taxonomies.

Prior to Darwin's time, classification was merely cataloging, and the cataloging system used morphology as the key characteristic for determining relationships. There is no theoretical basis for preferring one morphological cataloging scheme over another, however. The concept of evolution, and its prediction of common descent, provided that theoretical basis. Relationships, based on common ancestry, should provide a more accurate taxonomy. Besides being similar in size or shape, two organisms that were most closely related should have a common ancestor in the more recent past than would be the case for two less closely related organisms. The evolutionary history of the taxa was valuable and necessary information in this approach. The word for the determination of the evolutionary history of a species or group of species is **phylogenetics**, and the hypothesized evolutionary history and relationships of a species or group of species is a **phylogeny**. It quickly became clear that a taxonomic scheme that reflected phylogeny would be better than the arbitrary morphology-based schemes of the past. However, at the time of Darwin, and for many years thereafter, it was not exactly easy to discern the evolutionary history of organisms. So the development of a true phylogenetic taxonomy took a long time to develop, and, indeed, it is still being developed.

Taxonomy should reflect phylogeny

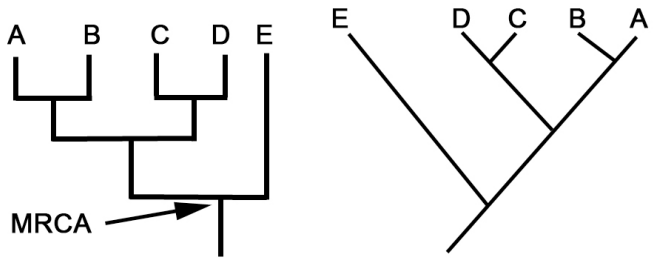


Figure 4.2 *A simple phylogenetic tree. MRCA = Most recent common ancestor*

The current approach to determining relationships between two (or more) groups of organisms is the construction of what are called **phylogenetic trees**. Phylogenetic trees are hypothesized reconstructions of evolutionary history. They depict arrays of extant (currently living) organisms at the tips of the branches, and branch points that indicate a common ancestor. (**Figure 4.2**). A, B, C, D, and E in this diagram are the species being considered. The vertical axis represents the passage of time. The branch points represent organisms that are putative common ancestors of the organisms on the branches above. In some cases these ancestors are known species from the fossil record. In most cases they are not. A common ancestor for all of these organisms is the branch point above the root of the tree, known as the **most recent common ancestor** (MRCA). These trees can be horizontal and vertical arrangements, or diagonal arrangements, both of which are shown here. The two arrangements, in this case, represent identical trees in terms of the hypothesized relationships of species A, B, C, D , and E.

Another critical point about these trees is that if you rotate the structures, using one of the branch points as a pivot, you don't change the relationships. So just like the two trees above, which show the same relationships even though they are formatted differently, all of the trees in Figure 4.3 (**Figure 4.3**) are essentially identical in terms of depicting the relationships between the species A, B, C, and D. If you don't see how that is true, just concentrate on the relationships and the branch points rather than on the sequence of species (A,B, C and D) shown across the tops of these diagrams. That sequence is not important; the branch structure underlying the sequence is what you will need to focus on.

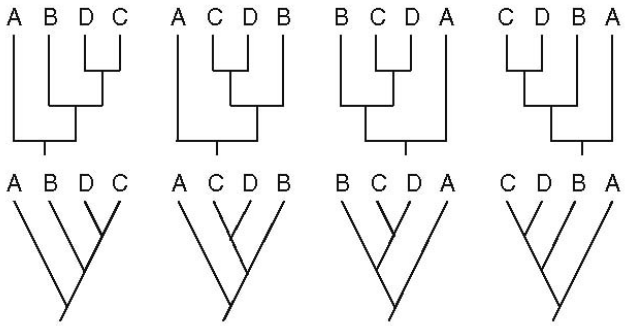


Figure 4.3 *Different depictions of the same phylogenetic tree. All of these depict the same relationships between the organisms A, B, C and D.*

To generate these tree structures, taxonomists use multiple characteristics to compare organisms, including external morphology, internal anatomy, behaviors, biochemical pathways, etc. Aristotle and Linneaus relied primarily on morphological characters (size and shape). However, as noted above, taxonomic schemes derived from comparing only a few characteristic can be very flawed. So modern taxonomists rely, if possible, on many characters, and taxonomies are constantly revised if additional characters are measured and included in the analysis. In addition, the development of a method called **cladistics** has revolutionized taxonomic thinking, and cladistics also depends on an understanding of evolutionary relationships. A **clade** is a group of organisms that consists of a common ancestor and all of its descendants. So another name for clade is a *monophyletic* group. For example, birds (including the bald eagle and your parakeet AND their ancestral organisms including one class of dinosaurs) are a clade. Clades can contain any number of species, but that number must include all of the descendants and the common ancestor. In the example above, A, B, C, D and E, as well as the ancestors indicated by the branch points, are a clade. A and B, plus their common ancestor, are also considered to be a clade.

What's the difference between a phylogeny, a phylogenetic tree, and a cladogram? For our purposes, there really isn't much of a difference. In this class we will use these terms interchangeably — they all describe a tree structure that represents the hypothesized evolutionary relationships within a group of organisms, based on data derived from various sources. The important things to remember are 1) organisms are related, and 2) that we can represent our hypotheses about these relationships with tree structures.

Cladistics relies on classifying characteristics of organisms as either **ancestral** or **derived** (another term that is sometimes used for ancestral characters is *primitive*). Ancestral characters are those inherited attributes that resemble those of an ancestor to the group. Derived characters are those features that are different from features found in the ancestor. The assumption of ancestral relationships in this approach is important, and signifies yet another way that evolutionary theory is the framework for understanding much of biology.

But how do you know which features are ancestral, and which are derived? There are several ways to do that, but for your purposes, you can assume that characters shared by all the organisms in the group are probably ancestral. Other characters, not shared by all the organisms in the group, are assumed to be derived. Lists of characters are generated, and used to prepare a tree-like structure known as a **cladogram**.

The best way to illustrate this process is with an example or two. Here is a table with some characters in rows and some organisms in columns. If there is an “X” in the column for an organism, the organism has that particular characteristic.

CHARACTER	Lamprey	Antelope	Bald Eagle	Alligator	Sea Bass
Lungs		X	X	X	
Jaws		X	X	X	X
Feathers			X		
Notochord	X	X	X	X	X
Gizzard			X	X	
Hair		X			

Figure 4.4 *List of characteristics of a group of animals that can be used to construct a cladogram.*

Which of these characters is common to all of these organisms? Obviously, it is the notochord. Assume that is the ancestral condition, and thus it should characterize the oldest ancestor at the base of any cladogram that you construct. Now you need to determine what the first branch in the tree is based on. That would be the character that is shared by all but one of these organisms, the presence of jaws. If you continue this process, adding a branch to the tree to depict groups that have, or do not have, a character, you will generate this cladogram (**Figure 4.5**).

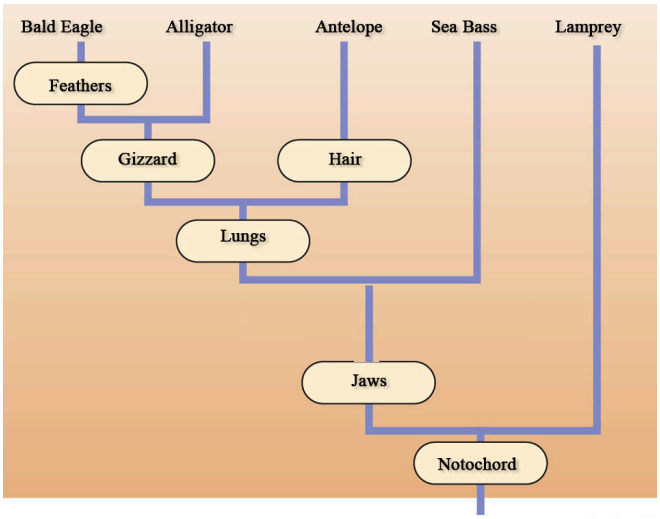


Figure 4.5 ***A cladogram constructed from the data in the table above.*** *Using the character list in the table above, and recognizing that the notochord is common to all of these organisms, a cladogram was constructed, using the other characteristics in that list.* Each shaded ellipse represents an ancestor possessing that trait, while the organism(s) on the opposite branch lack that trait.

Now that you know how to construct one, there are a few other things about cladograms that you need to know.

- The organisms along the top, or end, of a cladogram are not ordered in any particular way, and are not ancestors of each other. The order does not imply any sort of ranking from "primitive" to "advanced". As discussed in the previous chapter, evolution is not synonymous with progress. As discussed above, the branches, and not the sequence of organisms across the top, are the important information in these diagrams.
- The most closely related organisms have the most recent common ancestor, which is determined by counting the branch points backward from the organisms aligned across the top. For example , the bald Eagle and the Alligator have a more recent common ancestor than the Alligator and the Antelope; the Bald Eagle and the Alligator are more closely related than the Alligator and the Antelope.
- The lines between branches represents the arrow of time, but in most cases the length of these lines is not directly related to the length of time involved between the branch points. Some representations of cladograms include an axis where time can be deduced, but if there is no such axis, do not assume that a longer line represents a longer length of time.
- There are many ways to represent a cladogram, and many of these may look different, but will actually be identical. For example, the two cladograms shown in figure 4.2 depict the same set of relationships, even though one is vertically and horizontally aligned, and the other is diagonal in nature. In both cases organism C is more closely related to D (they share a relatively recent common ancestor) than A is related to B (which share a less recent common ancestor).

Cladograms can be generated from any set of inherited characters. But they become powerful tools, both for understanding the world and for generating predictive hypotheses, when they include characters that we know are quantitatively representative of relatedness, such as amino acid or nucleic acid sequences. But the single most important character is one that most directly reflects the degree of common ancestry – the DNA sequence. Since (as you will learn in detail later) DNA is the molecule of "descent with modification", similarities and differences in DNA sequences from different organisms are thus very useful in determining phylogenetic relationships. The development of DNA sequencing technology has allowed scientists to apply this tool to taxonomy, and this has resulted in an explosion of new and exciting discoveries about phylogeny.

Morphological traits, such as those used to construct the cladogram above, can be preserved in the fossil record, but DNA generally is not preserved. However, since we know that your DNA was inherited from your parents, and your grandparents, etc., we can predict that DNA sequences from related individuals would contain evidence of that relatedness. These sorts of "molecular fossils" can be used to produce and refine cladograms, generating the structures we know as phylogenetic trees.

"This view of life"

It is important to also understand that phylogenetic trees are **nested hierarchies**, i.e., any individual set of branches is also part of a larger set of branches. This is easily seen in the accompanying figure (**Figure 4.6**). The clade containing the eagle and the alligator nests within a clade that also contains the antelope, and that clade nests within another clade, and

another, etc. The hypothesis of descent with modification (i.e. evolution) absolutely predicts that the evolutionary history of organisms would be represented by a set of nested hierarchies. Cladistic analyses, particularly those based on DNA sequences, give results consistent with this prediction of Darwin's theory.

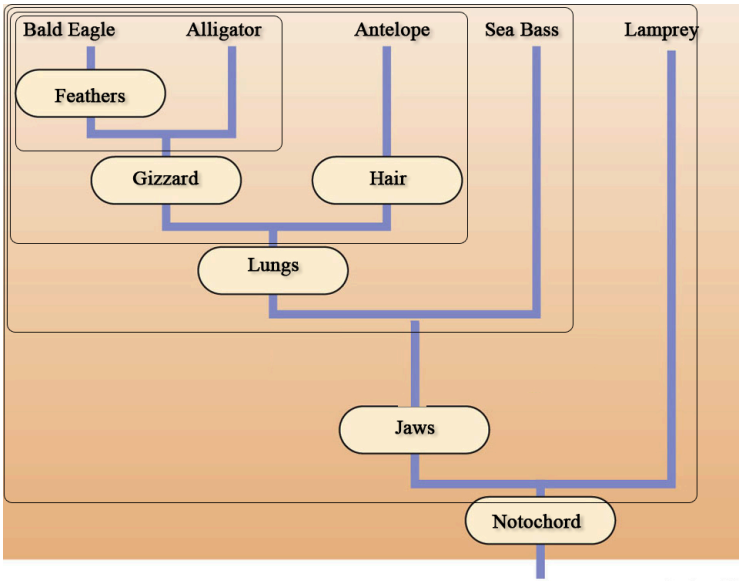


Figure 4.6 *Cladogram showing the nested hierarchical arrangement of organisms predicted by Darwin's theory of evolution.*

If we continue backward in time from any tip of any branch of the phylogenetic tree above, we soon realize that all of these clades have a common ancestor at some point in the distant past. DNA evidence supports and strengthens this interpretation. As a matter of fact, the observation that all life forms on the planet have DNA as their genetic material is further support of this interpretation. The details of the arrangements of the various branches may change, as new observations are included in the analyses of relationships. The figure below (**Figure 4.7**) shows various phylogenetic trees that have been used in the past. Note that the fundamental conclusion, that all life on this planet came from a common ancestor, is a constant feature of all of these trees. Indeed, as new observations are made, particularly in the realm of DNA sequence data, this conclusion has become increasingly well-supported.

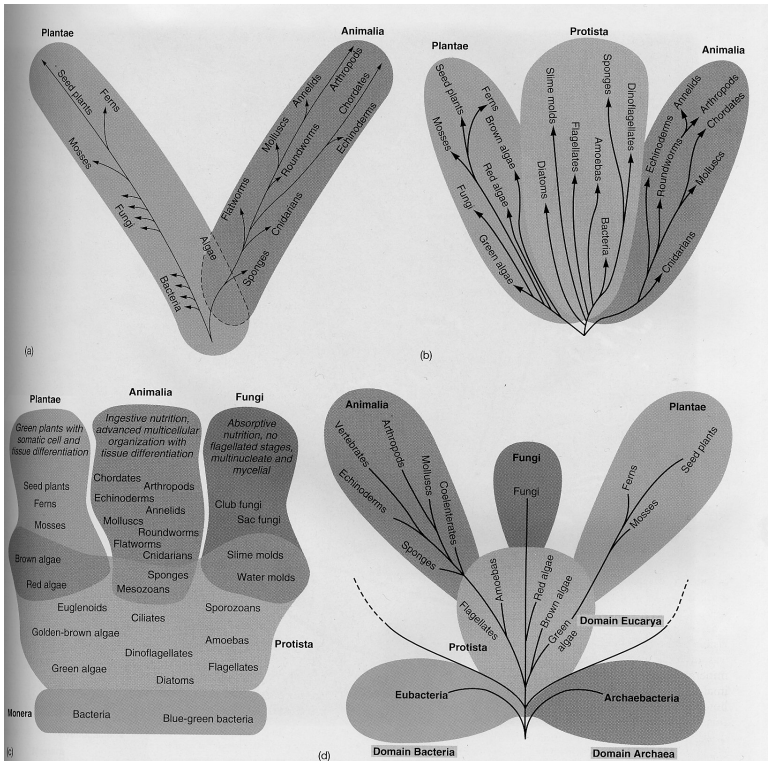


Figure 4.7 ***Various depictions of the Tree of Life.*** *These hypothesized trees reflect the state of knowledge at various times in the past. As new observations are made, new phylogenetic trees are constructed to accommodate that new knowledge. It is clear, from this example alone, that scientific hypotheses change in response to new discoveries and new knowledge. The currently accepted Tree of Life* ***Figure 4.8*** *will undoubtedly be modified in response to new scientific discoveries in the future.*

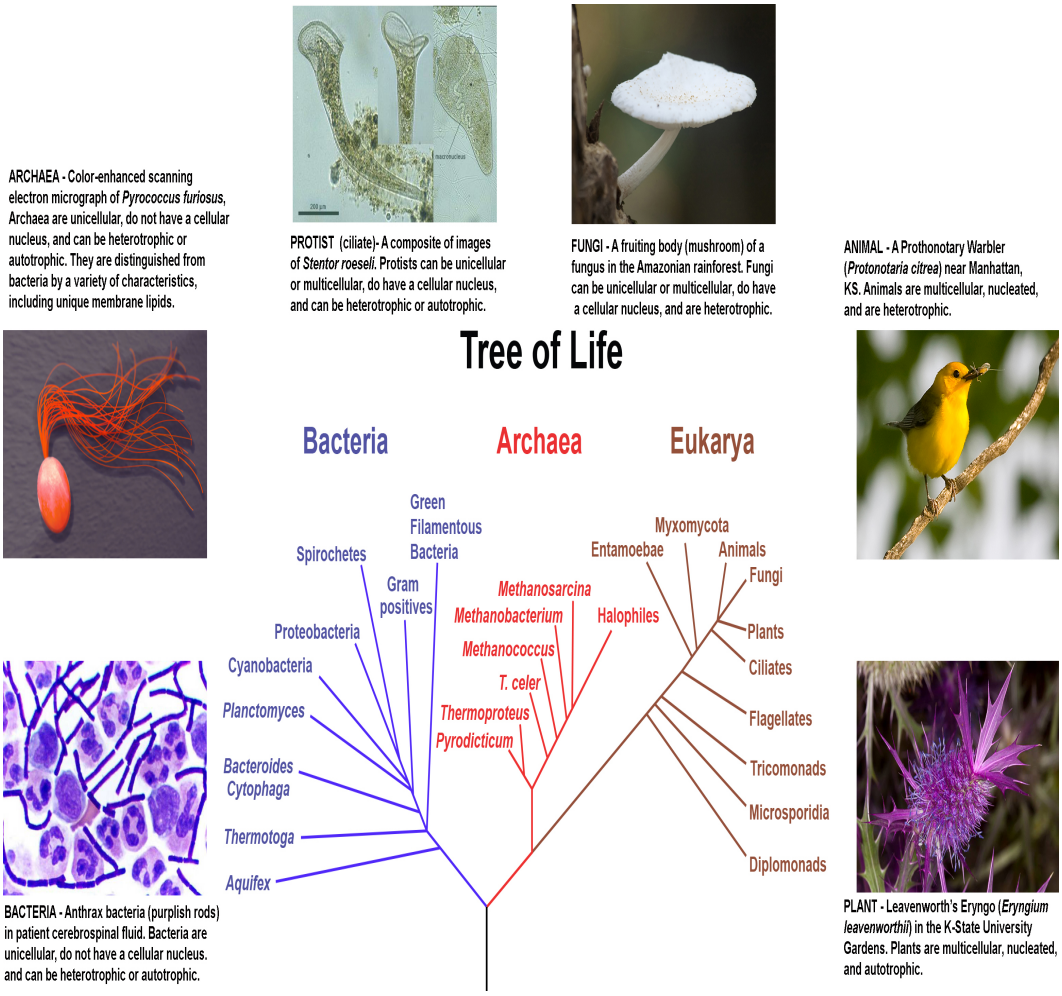


Figure 4.8 ***The Tree of Life.*** *The currently accepted hypothetical relationships between the three domains of living things. Bacteria, Archaea, and Eukarya (eukaryotes like you and me) are all thought to be descended from a common ancestor that lived billions of years ago.* Image credits: Anthrax bacteria, from John A. Jernigan et al, "Bioterrorism-Related Inhalational Anthrax: The First 10 Cases Reported in the United States". *Emerging Infectious Diseases* 7(6), December 2001. *Pyrococcus furiosus*, from Wikimedia Commons. By Fulvio314 (Own work) (http://creativecommons.org/licenses/by-sa/3.0) Stentor, copyright-free image from the Protist Image Database. Mushroom, warbler, and eryngo are used with permission of the photographer, David A. Rintoul.

The most widely accepted scheme has all living forms divided into three major Domains, **Bacteria**, **Archaea**, and **Eukarya**. The evolutionary relationship between these largest taxa is shown above (**Figure 4.8**). Note that current data support the idea that the Eukarya (including me and you) are more closely related to the Archaea than they are to the Bacteria; Archaea and Eukarya have a more recent common ancestor compared to Bacteria and Eukarya. Each of these Domains can be further subdivided into Kingdoms, Phyla, Classes, Orders, Families, Genera, and Species. For example, humans (*Homo sapiens*) are in the Domain Eukarya, Kingdom Animalia, Phylum Chordata, Class Mammalia, Order Primates, Family Hominidae, Genus *Homo*, and Species *sapiens*. In some cases these classifications are further organized into taxa such as Superfamily, or divided into taxa such as Suborder, but those are details that need not concern us here.

Indeed, the entire Tree of Life is a nested hierarchy. We're all related. Darwin suspected this, and his theory predicted the phylogenetic trees that scientists have generated from the observations and other data available to us. In one of the most often-quoted passages from the *Origin*, he demonstrates the sense of wonder that all biologists have when contemplating the diversity of living forms.

"There is grandeur in this view of life, with its several powers, having been originally breathed into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved." - Charles Darwin, 1859.

5 | INTRODUCTION TO ECOLOGY AND ECOSYSTEMS

5.1 | The Scope of Ecology

Introduction

> "Ecology is the entire science of the relations of the organism to the surrounding exterior world, to which relations we can count on the broader sense all of the conditions of existence. These are partly of organic, partly of inorganic nature... It is the "household' of nature."

- Ernst Haeckel, 1869, inventing and defining the word "ecology".

Ecology is the study of the interactions of living organisms with their environment. The Greek root of the word, and the basis for Haeckel's analogy above, is οἶκος (oikos), which means house or dwelling-place. Ecology simply means "knowledge of the house". One core goal of ecology is to understand the distribution and abundance of living things in the physical environment. Attainment of this goal requires the integration of scientific disciplines inside and outside of biology, such as biochemistry, physiology, evolution, biodiversity, molecular biology, geology, and climatology. Some ecological research also applies aspects of chemistry and physics, and it frequently uses mathematical models.

Levels of Ecological Study

When a discipline such as biology is studied, it is often helpful to subdivide it into smaller, related areas. For instance, cell biologists interested in cell signaling need to understand the chemistry of the signaling molecules (which can be sugars, proteins, lipids, gases, or other compounds), as well as the result of cell signaling. Ecologists interested in the factors that influence the survival of an endangered species might use mathematical models to predict how current conservation efforts affect endangered organisms. To produce a sound set of management options, a conservation biologist needs to collect accurate data, including current population size, factors affecting reproduction (like physiology and behavior), habitat requirements (such as plants and soils), and potential human influences on the endangered population and its habitat (which might be derived through studies in sociology and urban ecology). Within the discipline of ecology, researchers work at four specific levels, sometimes discretely and sometimes with overlap: organism, population, community, and ecosystem (**Figure 5.1**).

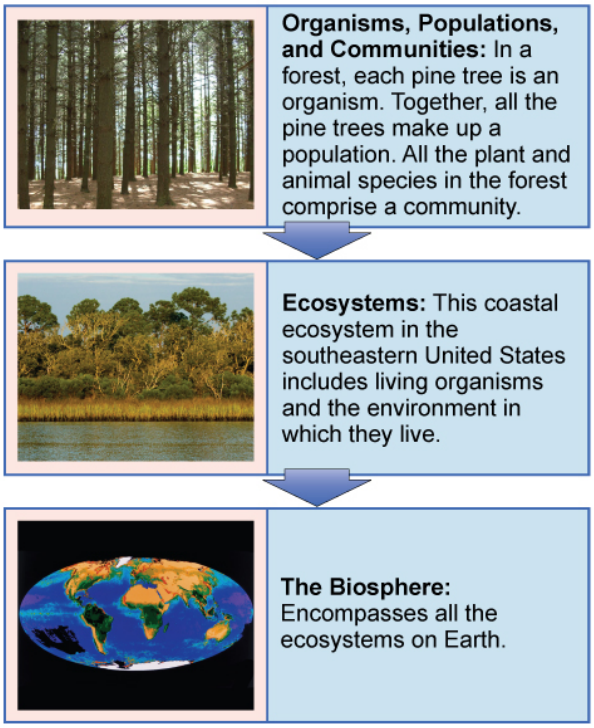


Figure 5.1 Ecologists study within several biological levels of organization. (credit “organisms”: modification of work by "Crystl"/Flickr; credit “ecosystems”: modification of work by Tom Carlisle, US Fish and Wildlife Service Headquarters; credit “biosphere”: NASA)

Organismal Ecology

Researchers studying ecology at the organismal level are interested in the adaptations that enable individuals to live in specific habitats. These adaptations can be morphological, physiological, and behavioral. For instance, the Karner blue butterfly (*Lycaeides melissa samuelis*) (**Figure 5.2**) is considered a specialist because the females preferentially oviposit (that is, lay eggs) on wild lupine. This preferential adaptation means that the Karner blue butterfly is highly dependent on the presence of wild lupine plants for its continued survival.

Figure 5.2 The Karner blue butterfly (*Lycaeides melissa samuelis*) is a rare butterfly that lives only in open areas with few trees or shrubs, such as pine barrens and oak savannas. It can only lay its eggs on lupine plants. (credit: modification of work by J & K Hollingsworth, USFWS)

After hatching, the larval caterpillars emerge and spend four to six weeks feeding solely on wild lupine (**Figure 5.3**). The caterpillars pupate (undergo metamorphosis) and emerge as butterflies after about four weeks. The adult butterflies feed on the nectar of flowers of wild lupine and other plant species. A researcher interested in studying Karner blue butterflies at the organismal level might, in addition to asking questions about egg laying, also ask questions about the butterflies’ preferred temperature (a physiological question), or the behavior of the caterpillars when they are at different larval stages (a behavioral question).

Figure 5.3 The wild lupine (*Lupinus perennis*) is the host plant for the Karner blue butterfly.

Population Ecology

A **population** is a group of interbreeding organisms that are members of the same species living in the same area at the same time. A population is identified, in part, by where it lives, and the area it occupies, which may have natural or artificial boundaries. Natural boundaries might be rivers, mountains, or deserts, while examples of artificial boundaries include mowed grass, man-made structures, or roads. The study of **population ecology** focuses on the number of individuals in an area and asks how and why population size changes over time. Population ecologists are particularly interested in counting the Karner blue butterfly, for example, because it is classified as federally endangered. However, the distribution and density of this species is highly influenced by the distribution and abundance of wild lupine. Researchers might ask questions about the factors leading to the decline of wild lupine and how these affect Karner blue butterflies. For example, ecologists know that wild lupine thrives in open areas where trees and shrubs are largely absent. In natural settings, intermittent wildfires regularly remove trees and shrubs, helping to maintain the open areas that wild lupine requires. Mathematical models can be used to understand how wildfire suppression by humans has led to the decline of this important plant for the Karner blue butterfly.

Community Ecology

A biological **community** consists of the populations of all the species within an area, typically a three-dimensional space, and the interactions within and among these species. Community ecologists are interested in the processes driving these interactions and their consequences. Questions about **intraspecific interactions** (interactions between members of the same species) often focus on **competition** among members of the same species for a limited resource. Ecologists also study interactions between members of different species; these are called **interspecific interactions**. Examples of interspecific interactions include predation, parasitism, herbivory, competition, and pollination. These interactions can have regulating effects on population sizes, and can impact the ecological and evolutionary processes, eventually affecting species diversity.

For example, Karner blue butterfly larvae form mutualistic relationships with ants. **Mutualism** is a form of long-term relationship that has coevolved between two species and from which each species benefits. For mutualism to exist between individual organisms, each species must receive some benefit from the other as a consequence of the relationship. Researchers have shown that there is an increase in the probability of survival when Karner blue butterfly larvae (caterpillars) are tended by ants. This might be because the larvae spend less time in each life stage when tended by ants, which provides an advantage for the larvae. Meanwhile, the Karner blue butterfly larvae secrete a carbohydrate-rich substance that is an important energy source for the ants. Both the Karner blue larvae and the ants benefit from their interaction.

Ecosystem Ecology

Ecosystem ecology is an extension of organismal, population, and community ecology. The **ecosystem** is composed of all

the **biotic** components (living things) in an area along with the **abiotic** components (non-living things) of that area; this definition hearkens back to Haeckel's original definition (above). Some of the abiotic components include air, water, and soil. Ecosystem biologists ask questions about how nutrients and energy are stored and how they move among organisms and the surrounding atmosphere, soil, and water.

The Karner blue butterflies and wild lupine live in an oak-pine barren habitat. This habitat is characterized by natural disturbance and nutrient-poor soils that are low in nitrogen. The availability of nutrients is an important factor in the distribution of the plants that live in this habitat. Researchers interested in ecosystem ecology could ask questions about the importance of limited resources and the movement of resources, such as nutrients, through the biotic and abiotic portions of the ecosystem.

career CONNECTION

Ecologist

A career in ecology contributes to many facets of human society. Ecologists can conduct their research in the laboratory and outside in natural environments (**Figure 5.4**). These natural environments can be as close to home as the stream running through your campus or as far away as the hydrothermal vents at the bottom of the Pacific Ocean. Ecologists manage natural resources such as white-tailed deer populations (*Odocoileus virginianus*) for hunting or aspen (*Populus* spp.) timber stands for paper production. Ecologists also work as educators who teach children and adults at various institutions including universities, high schools, museums, and nature centers. Ecologists may also work in advisory positions assisting local, state, and federal policymakers to develop laws that are ecologically sound, or they may develop those policies and legislation themselves. To become an ecologist requires an undergraduate degree, usually in a natural science. The undergraduate degree is often followed by specialized training or an advanced degree, depending on the area of ecology selected. Ecologists should also have a broad background in the physical sciences, as well as a sound foundation in mathematics and statistics. But even if you don't plan to become an ecologist, understanding ecological issues as a voter and citizen can help society meet the basic human needs of food, shelter, and health care.

Figure 5.4 This landscape ecologist is releasing a black-footed ferret (shown at right) into its native habitat as part of a study. (credit - biologist photo: USFWS Mountain Prairie Region, NPS; black-footed ferret photo, David A. Rintoul)

5.2 | Ecology of Ecosystems

Introduction

> "The early bird gets the worm, but the second mouse gets the cheese."

Willie Nelson, American musician

Competition for limited resources, whether it is a worm or a chunk of cheese, is an essential component of the evolutionary mechanism we call natural selection. Competition in communities (all living things within specific habitats) is observed both between members of the same species, and between members of different species. The resources for which organisms compete include food (or sunlight in the case of plants), mineral nutrients, nesting habitat, etc. Other critical factors influencing community dynamics are the components of its physical and geographic environment: a habitat's latitude, amount of rainfall, topography (elevation), and temperature. These are all important environmental variables that help determine which organisms can exist within a particular area.

An **ecosystem** is a community of living organisms and their interactions with their abiotic (non-living) environment. Ecosystems can be small, such as the tide pools found near the rocky shores of many oceans, or large, such as the Amazon Rainforest in Brazil (**Figure 5.5**).

(a) (b)

Figure 5.5 A (a) tidal pool ecosystem in Matinicus Island in Maine is a small ecosystem, while the (b) Amazon Rainforest in Brazil is a large ecosystem. (credit a: modification of work by "takomabibelot"/Flickr; credit b: modification of work by Ivan Mlinaric)

There are three broad categories of ecosystems based on their general environment: freshwater, ocean water, and terrestrial. Within these broad categories are individual ecosystem types based on the organisms present and the type of environmental habitat.

Ocean ecosystems are the most common, covering 75 percent of the Earth's surface and consisting of two general types: shallow ocean (near islands and continents), and the deep ocean. The shallow ocean ecosystems include extremely biodiverse coral reef ecosystems. The surface water of the deep ocean is known for its large numbers of plankton (small organisms that disperse at the mercy of the winds and currents). Planktonic organisms can be phytoplankton (photosynthetic organisms), or zooplankton (tiny animals or animal larvae, which feed on the phytoplankton). These two environments are globally important; the phytoplankton perform 40 percent of all photosynthesis on the planet (i.e., produce 40% of the oxygen and fix 40% of the CO_2). Although not as diverse as the other two, deep ocean ecosystems contain a wide variety of marine organisms. Such ecosystems exist even at the bottom of the ocean where light is unable to penetrate.

Freshwater ecosystems are the rarest, occurring on only 1.8 percent of the Earth's surface. Lakes, rivers, streams, and springs comprise these systems; they are quite diverse, and support a variety of fish, amphibians, reptiles, insects, phytoplankton, fungi, and bacteria.

Terrestrial ecosystems, also known for their diversity, are grouped into large categories called biomes, such as tropical rain forests, savannas, deserts, coniferous forests, deciduous forests, and tundra. Grouping these ecosystems into just a few biome categories obscures the great diversity of the individual ecosystems within them. For example, there is great variation in desert biomes: the saguaro cacti and other plant life in the Sonoran Desert, in the United States and Mexico, are relatively abundant compared to the lack of plant life in the desolate rocky desert of Boa Vista, an island off the coast of Western Africa (**Figure 5.6**).

(a) (b)

Figure 5.6 Desert ecosystems, like all ecosystems, can vary greatly. The desert in (a) Saguaro National Park, Arizona, has abundant plant life, while the rocky desert of (b) Boa Vista island, Cape Verde, Africa, is devoid of plant life. (credit a: modification of work by Jay Galvin; credit b: modification of work by Ingo Wölbern)

Ecosystems are complex with many interacting parts, and dissecting the roles of these interacting components can be a challenge. Furthermore, ecosystems are routinely exposed to various disturbances, or changes in the environment that affect their compositions. For example, variations in rainfall and temperature can affect patterns and rates of plant growth, even though this may take several years. Many disturbances are a result of natural processes. For example, fire is a disturbance that can be caused by a lightning strike in a prairie or forest ecosystem. Recovery from disturbances can be highly variable as well; some ecosystems (e.g. prairie) are adapted to fire and can regenerate quickly. Others might recover more slowly. Other disturbances are the result of human activities. The impact of environmental disturbances caused by human activities is as important as the changes wrought by natural processes. Human agricultural practices, air pollution, acid rain, global deforestation, overfishing, eutrophication, oil spills, and illegal dumping on land and into the ocean are disturbances, and biological responses to these are of interest to ecologists and conservationists.

Equilibrium is the steady state of an ecosystem where all organisms are in balance with their environment and with each other. In ecology, two parameters are used to measure changes in ecosystems: resistance and resilience. The ability of an ecosystem to remain at equilibrium in spite of disturbances is called resistance. The speed at which an ecosystem recovers equilibrium after being disturbed is its resilience. Ecosystem resistance and resilience are especially important when considering human impact. The nature of an ecosystem may change to such a degree that it can lose its resilience entirely. This process can lead to the complete destruction or irreversible altering of the ecosystem.

Food Chains and Food Webs

The term “food chain” is sometimes used metaphorically to describe human social situations. In this sense, food chains are thought of as a competition for survival, such as “who eats whom?” Someone eats and someone is eaten. Therefore, it is not surprising that in our competitive “dog-eat-dog” society, individuals who are considered successful are seen as being "at the top of the food chain", consuming all others for their benefit, whereas the less successful are seen as being at the bottom.

The scientific understanding of a food chain is more precise than in its everyday usage. In ecology, a **food chain** is a linear sequence of organisms through which nutrients and energy pass: primary producers, primary consumers, and higher-level consumers are used to describe ecosystem structure and dynamics. There is a single path through the chain. Each organism in a food chain occupies what is called a **trophic level** (composed of organisms that share the same function in the food chain and the same nutritional relationship to the primary sources of energy). Depending on their role as producers or consumers, species or groups of species can be assigned to one or more trophic levels; for example bears eat plants (and thus are primary consumers) and also eat other animals (and thus are secondary or tertiary consumers).

In many ecosystems, the bottom of the food chain consists of photosynthetic organisms (plants and/or phytoplankton), which are called **primary producers**. The organisms that consume the primary producers are herbivores: the **primary consumers**. **Secondary consumers** are usually carnivores that eat the primary consumers. **Tertiary consumers** are carnivores that eat other carnivores. Higher-level consumers feed on the next lower tropic levels, and so on, up to the organisms at the top of the food chain: the **apex consumers**. In the Lake Ontario food chain shown in **Figure 5.7**, the Chinook salmon is the apex consumer at the top of this food chain.

One of the classes of consumers deserves special mention; these are the **decomposers**, which break down waste or dead organic matter. Fungi and bacteria are decomposers in many ecosystems, utilizing the chemical energy in dead organic material to fuel their own metabolic processes. Some of the decomposers are also known as **detritivores** (literally, detritus- or debris-eaters). These are generally multicellular animals such as earthworms, crabs, slugs, vultures, etc. which not only feed on dead organic matter, but often fragment it as well, making it more available for bacterial or fungal decomposers.

These organisms have a critical role in ecosystems, and are one of the main mechanisms by which nutrients get recycled for other organisms to use again. Even apex consumers such as the Chinook salmon have to die sometime, and the nutrients in their bodies can nourish a host of detritivores and decomposers.

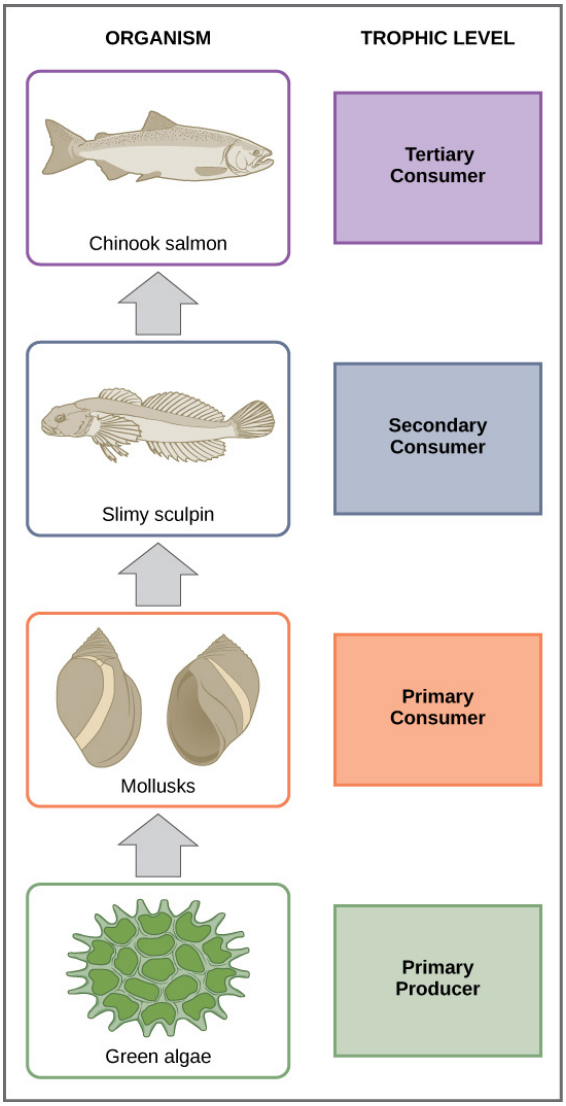


Figure 5.7 These are the trophic levels of a food chain in Lake Ontario at the United States-Canada border. Energy and nutrients flow from photosynthetic green algae at the bottom to the top of the food chain: the introduced Chinook salmon.

One major factor that limits the length of food chains is energy. Approximately 90% of the energy entering the system (from sunlight converted to carbohydrates by the primary producers) is lost as heat between one trophic level to the next, as explained by the second law of thermodynamics. Put another way, only 10% of the energy in a lower trophic level is transferred to the next trophic level. Thus, after a limited number of trophic levels (energy transfers), the amount of energy remaining in the food chain may not be great enough to support viable populations at yet a higher trophic level.

The loss of energy between trophic levels is illustrated by the pioneering studies of Howard T. Odum in the Silver Springs, Florida, ecosystem in the 1940s (**Figure 5.8**). The primary producers contained 20,819 kcal/m^2/yr (kilocalories per square meter per year), the primary consumers contained 3368 kcal/m^2/yr, the secondary consumers contained 383 kcal/m^2/yr, and the tertiary consumers only conained 21 kcal/m^2/yr. Thus, there is little energy remaining for another level of consumers in this ecosystem.

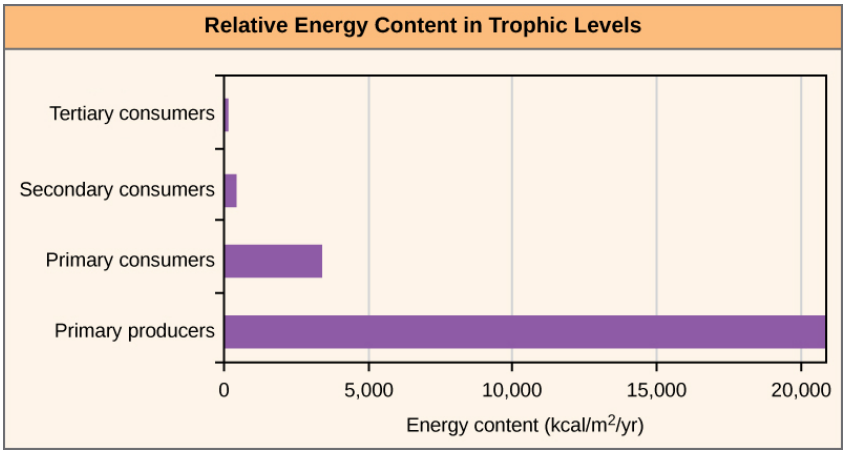


Figure 5.8 The relative energy in trophic levels in a Silver Springs, Florida, ecosystem is shown. Each trophic level has less energy available and supports fewer organisms at the next level.

There is a one problem when using food chains to describe most ecosystems. Even when all organisms are grouped into appropriate trophic levels, some of these organisms can feed on species from more than one trophic level, as noted above. Likewise, some of these organisms can be eaten by species from multiple trophic levels. In other words, a strictly linear model of ecosystems, the food chain, does not completely describe ecosystem structure. A holistic model — which accounts for all the interactions between different species and their complex interconnected relationships with each other and with the environment — is a more accurate and descriptive model for ecosystems. We call such models food webs. A **food web** is a graphic representation of a holistic, non-linear web of primary producers, primary consumers, and higher-level consumers used to describe ecosystem structure and dynamics (**Figure 5.9**).

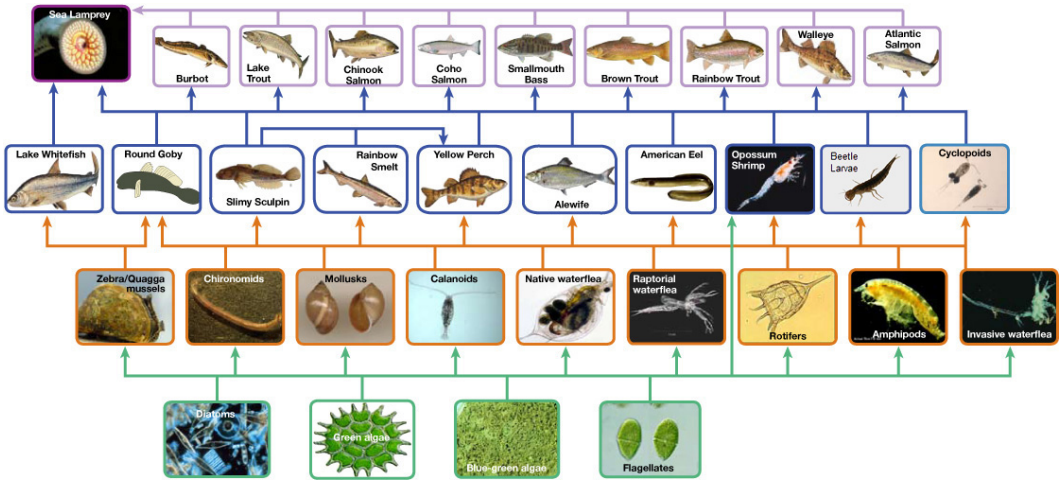


Figure 5.9 This food web shows the interactions between organisms across trophic levels in the Lake Ontario ecosystem. Primary producers are outlined in green, primary consumers in orange, secondary consumers in blue, and tertiary consumers in purple. Arrows point from an organism that is consumed to the organism that consumes it. Notice how some lines point to more than one trophic level. For example, the opossum shrimp eats both primary producers and primary consumers. (credit: NOAA, GLERL)

A comparison of the two types of structural ecosystem models reveals strengths for both models. Food chains are more flexible for analytical modeling, are easier to follow, and are easier to experiment with, whereas food web models more accurately represent ecosystem structure and dynamics, and data can be directly used as input for simulation modeling.

Two general types of food webs are often shown interacting within a single ecosystem. A grazing food web (such as the Lake Ontario food web in **Figure 5.9**) has plants or other photosynthetic organisms at its base, followed by herbivores and various carnivores. A detrital food web consists of a base of organisms that feed on decaying organic matter (dead organisms), called decomposers or detritivores. These organisms are usually bacteria or fungi that recycle organic material back into the biotic part of the ecosystem as they themselves are consumed by other organisms. As all ecosystems require a method to recycle material from dead organisms, most grazing food webs have an associated detrital food web. For example, in a meadow ecosystem, plants may support a grazing food web of different organisms, primary and other levels

of consumers, while at the same time supporting a detrital food web of bacteria, fungi, and detrivorous invertebrates feeding off dead plants and animals.

evolution CONNECTION

Three-spined Stickleback

It is well established that changes in the environment play a major role in the evolution of species within an ecosystem. However, little is known about how the evolution of species within an ecosystem can alter the ecosystem environment. In 2009, Dr. Luke Harmon, from the University of Idaho in Moscow, published a paper that for the first time showed that the evolution of organisms into subspecies can have direct effects on their ecosystem environment.[1]

The three-spine stickleback (*Gasterosteus aculeatus*) is a freshwater fish that evolved from a saltwater fish. Evolutionary changes enabled it to live in freshwater lakes about 10,000 years ago, which is considered a recent development in evolutionary time (**Figure 5.10**). Over the last 10,000 years, these freshwater fish then became isolated from each other in different lakes. Depending on which lake population was studied, findings showed that these sticklebacks then either remained as one species or evolved into two species. The divergence of species occurred because different populations used different areas of the lake for feeding.

Dr. Harmon and his team created artificial pond microcosms in 250-gallon tanks and added muck from freshwater ponds as a source of zooplankton and other invertebrates to sustain the fish. In different experimental tanks they introduced one species of stickleback from either a single-species or double-species lake.

Over time, the team observed that some of the tanks bloomed with algae while others did not. This puzzled the scientists, and they decided to measure some water quality parameters, including the amount of dissolved organic carbon (DOC). DOC consists of organic compounds such as amino acids, carbohydrates, lignins, and many others; these are usually derived from decomposition of plant material in the water. DOC can vary not only in composition, but also in the size of the particles. Larger aggregates of decaying organic matter can give pond-water its slightly brownish color. It turned out that the water from the tanks with two-species fish contained larger particles of DOC (and hence darker water) than water with single-species fish. This increase in DOC blocked the sunlight and prevented algal blooming. Conversely, the water from the single-species tank contained smaller DOC particles, allowing more sunlight penetration to fuel the algal blooms. As the authors point out, "sticklebacks act as ecosystem engineers, strongly affecting the composition of the DOC pool and the physical light environment."

This change in the environment, which is due to the different feeding habits of the stickleback species in each lake type, probably has a great impact on the survival of other species in these ecosystems, especially other photosynthetic organisms. Thus, the study shows that, at least in these ecosystems, the environment and the evolution of populations have reciprocal effects that may now be factored into simulation models.

Figure 5.10 The three-spined stickleback evolved from a saltwater fish to freshwater fish. (credit: Barrett Paul, USFWS)

5.3 | The Laws of Thermodynamics

Introduction

> "Nothing in life is certain except death, taxes and the second law of thermodynamics. All three are processes in which useful or accessible forms of some quantity, such as energy or money, are transformed into useless, inaccessible forms of the same quantity. That is not to say that these three processes don't have fringe benefits: taxes pay for roads and schools; the second law of thermodynamics drives cars, computers and metabolism; and death, at the very least, opens up tenured faculty positions."

Seth Lloyd, *Nature* 430, 971 (26 August 2004)

Thermodynamics refers to the study of energy and energy transfer involving physical matter. The matter and its environment relevant to a particular case of energy transfer are classified as a system, and everything outside of that system 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. An **open system** is one in which energy can be transferred between the system and its surroundings. The stovetop system is open because heat can be lost into the air. A **closed system** is one that cannot transfer energy to its surroundings.

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

The First Law of Thermodynamics

The first law of thermodynamics deals with the total amount of energy in a system. It states that the total amount of energy in a closed system is constant. 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 within a system, but it cannot be created or destroyed. This is the principle of conservation of energy. Transfers and transformations of energy take place around us all the time. Light bulbs transform electrical energy into light 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 into the chemical energy stored within organic molecules. Some examples of energy transformations are shown in **Figure 5.11**.

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 very well. Chemical energy stored within organic molecules such as sugars and fats is 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 performed by cells include building complex molecules, transporting materials, powering the beating motion of cilia or flagella, contracting muscle fibers to create movement, and reproduction.

1. *Nature* (Vol. 458, April 1, 2009)

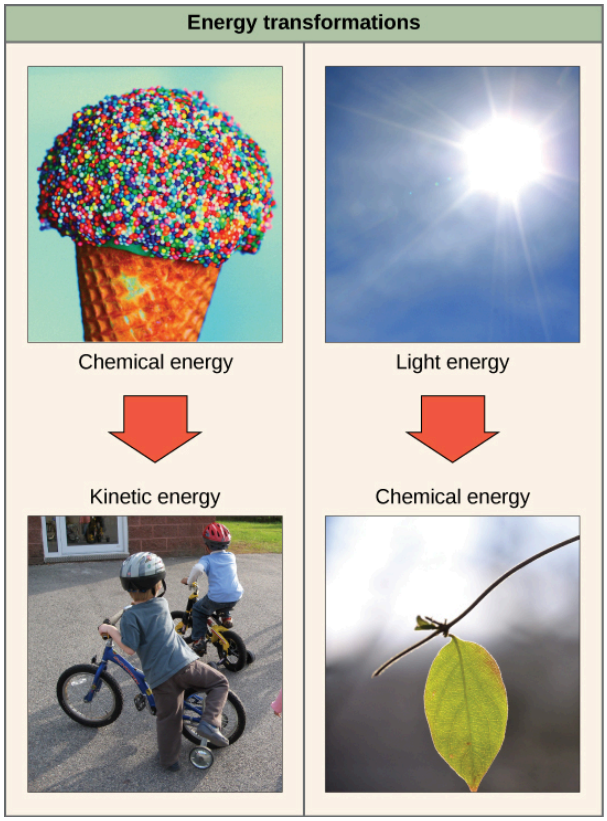


Figure 5.11 Shown are two examples of energy being transferred from one system to another and transformed from one form to another. Humans can convert the chemical energy in food, like this ice cream cone, into kinetic energy (the energy of movement to ride a bicycle). Plants can convert electromagnetic radiation (light energy) from the sun into chemical energy. (credit “ice cream”: modification of work by D. Sharon Pruitt; credit “kids on bikes”: modification of work by Michelle Riggen-Ransom; credit “leaf”: modification of work by Cory Zanker)

The Second Law of Thermodynamics

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. None of the energy transfers we’ve discussed is completely 100 percent 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 doing work. For example, when an airplane flies through the air, some of the energy of the flying plane is lost as heat energy due to friction with the surrounding air. This friction actually heats the air by temporarily increasing the speed of air molecules. Likewise, some energy is lost as heat energy during cellular metabolic reactions. This is good for warm-blooded creatures like us, because heat energy helps to maintain our body temperature. Strictly speaking, no energy transfer is completely efficient, because some energy is always lost in an unusable form.

An important concept in physical systems is that of order and disorder (also known as randomness). 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 is a state with high disorder and low energy. To better understand entropy, think of a student’s bedroom. If no energy or work were put into it, the room would quickly become messy. It would exist in a very disordered state, one of high entropy. Energy must be put into the system, in the form of the student doing work and putting everything away, in order to bring the room back to a state of cleanliness and order. This state is one of low entropy. Similarly, a car or house must be constantly maintained with work in order to keep it in an ordered state. Left alone, the entropy of the house or car gradually increases through rust and degradation. Molecules and chemical reactions have varying amounts of entropy as well. For example, as chemical reactions reach a state of equilibrium, entropy increases, and as molecules at a high concentration in one place diffuse and spread out, entropy also increases.

5.4 | Energy Flow

> “All flesh is grass.”

- Isaiah 40:6

All living things require energy in one form or another. Energy is required by most complex metabolic pathways (often in the form of adenosine triphosphate, ATP), especially those responsible for building large molecules from smaller compounds, and life itself is an energy-driven process. Living organisms would not be able to assemble macromolecules (proteins, lipids, nucleic acids, and complex carbohydrates) without a constant energy input. And yes, Isaiah was correct; your flesh and all flesh originated in the green plants of this planet.

It is important to understand how organisms acquire energy and how that energy is passed from one organism to another through food webs and their constituent food chains. Food webs illustrate how energy flows directionally through ecosystems, including how efficiently organisms acquire it, use it, and how much remains for use by other organisms of the food web. The transfer of energy from one trophic level to the next is an important concept when examining the effects of human disturbance on an ecosystem.

How Organisms Acquire Energy in a Food Web

Energy is acquired by living things in three ways: **photosynthesis**, **chemosynthesis**, and the consumption and digestion of other living or previously living organisms. Organisms using the first two methods are called **autotrophs** (auto=self, these organisms feed themselves). Those that rely on consumption of others are called **heterotrophs** (hetero=other, these organisms are fed by others). Photosynthetic autotrophs (photoautotrophs) use sunlight as an energy source, whereas chemosynthetic autotrophs (chemoautotrophs) use inorganic molecules as an energy source. Autotrophs are critical for all ecosystems. Without these organisms, energy would not be available to other living organisms and life itself would not be possible.

Photoautotrophs, such as plants, algae, and photosynthetic bacteria, serve as the energy source for a majority of the world’s ecosystems. These ecosystems are often described by grazing food webs. Photoautotrophs harness the solar energy of the sun by converting it to chemical energy in the form of ATP (and NADP). The energy stored in ATP is used to synthesize complex organic molecules, such as glucose.

Chemoautotrophs are primarily bacteria that are found in rare ecosystems where sunlight is not available, such as in those associated with dark caves or hydrothermal vents at the bottom of the ocean (**Figure 5.12**). Many chemoautotrophs in hydrothermal vents use hydrogen sulfide (H_2S) released from the vents as a source of chemical energy. This allows chemoautotrophs to synthesize complex organic molecules, such as glucose, for their own energy and in turn supplies energy to the rest of the ecosystem, making chemoautotrophs the primary producers of their ecosystem.

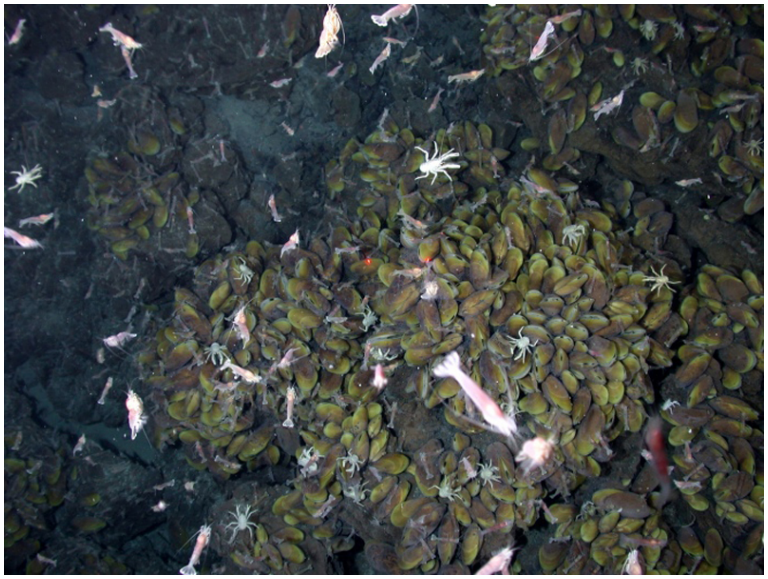


Figure 5.12 Swimming shrimp, a few squat lobsters, and hundreds of vent mussels are seen at a hydrothermal vent at the bottom of the ocean. As no sunlight penetrates to this depth, the ecosystem is supported by chemoautotrophic bacteria and organic material that sinks from the ocean’s surface. This picture was taken in 2006 at the submerged NW Eifuku volcano off the coast of Japan by the National Oceanic and Atmospheric Administration (NOAA). The summit of this highly active volcano lies 1535 m below the surface.

Productivity within Trophic Levels

The general term "production" refers to the amount of new organic matter generated by photosynthesis. Biologists have found it useful to define and measure production more precisely. The total amount of solar energy captured by photosynthesis is **gross primary productivity** or GPP. Producers can use some of this energy for their own maintenance and metabolism, and also use some to produce organic compounds that they accumulate. This latter fraction is called **net primary productivity** or NPP. This is the only energy, contained in the **biomass** of the producers, that is available to all the heterotrophic organisms in that ecosystem.

Studies on both natural and agricultural systems have shown that plants capture and convert about 1.3 - 1.6% of the solar energy that is available. They use about a quarter of that captured energy for their own metabolism and maintenance, leaving only about 1% of the solar energy as net primary productivity. The standard unit for measuring productivity is grams of biomass per square meter per year, but it is important to remember that this biomass is actually converted to energy by the heterotrophic organisms. Productivity in various ecosystems ranges from approximately 2,000 g/m^2/yr in tropical forests, salt marches, etc., to less than 100 g/m^2/yr in some desert ecosystems. An example of gross primary productivity is shown in the compartment diagram of energy flow within the Silver Springs aquatic ecosystem (**Figure 5.13**). In this ecosystem, the total energy accumulated by the primary producers (gross primary productivity) was shown to be 20,810 kcal/m^2/yr.

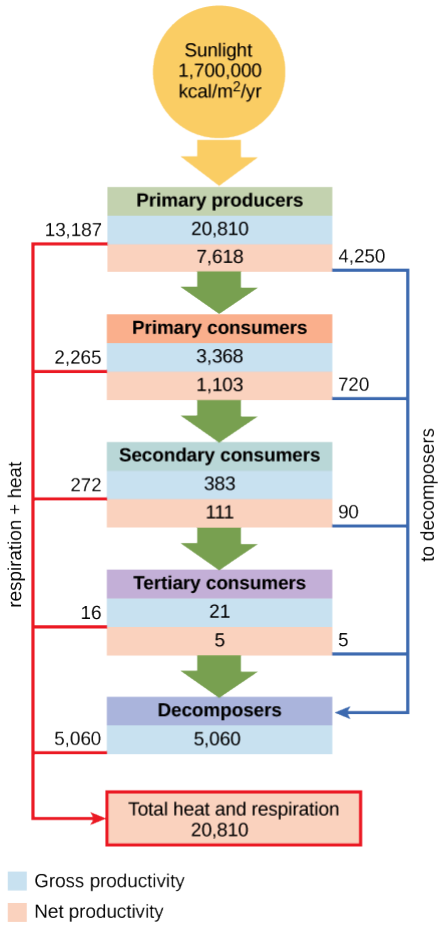


Figure 5.13 Trophic levels in an aquatic ecosystem Trophic levels and energy flow in the Silver Spring ecosystem

Modeling Ecosystems Energy Flow: Ecological Pyramids

The structure of ecosystems can be visualized with ecological pyramids, which were first described by the pioneering studies of Charles Elton in the 1920s. **Ecological pyramids** show the relative amounts of various parameters (such as number of organisms, energy, and biomass) across trophic levels.

Pyramids of numbers can be either upright or inverted, depending on the ecosystem. As shown in **Figure 5.14**, typical grassland during the summer has a base of many plants and the numbers of organisms decrease at each trophic level.

However, during the summer in a temperate forest, the base of the pyramid consists of few trees compared with the number of primary consumers, mostly insects. Because trees are large, they have great photosynthetic capability, and dominate other plants in this ecosystem to obtain sunlight. Even in smaller numbers, primary producers in forests are still capable of supporting other trophic levels.

Another way to visualize ecosystem structure is with pyramids of biomass. This pyramid measures the amount of energy converted into living tissue at the different trophic levels. Using the Silver Springs ecosystem example, these data exhibit an upright biomass pyramid (**Figure 5.14**), whereas the pyramid from the English Channel example is inverted. The plants (primary producers) of the Silver Springs ecosystem make up a large percentage of the biomass found there. However, the phytoplankton in the English Channel example make up less biomass than the primary consumers, the zooplankton. As with inverted pyramids of numbers, this inverted pyramid is not due to a lack of productivity from the primary producers, but results from the high turnover rate of the phytoplankton. The phytoplankton are consumed rapidly by the primary consumers, thus, minimizing their biomass at any particular point in time. However, phytoplankton reproduce quickly, thus they are able to support the rest of the ecosystem.

Pyramid ecosystem modeling can also be used to show energy flow through the trophic levels. Notice that these numbers are the same as those used in the energy flow compartment diagram in (**Figure 5.13**). Pyramids of energy are always upright, and an ecosystem without sufficient primary productivity cannot be supported. All types of ecological pyramids are useful for characterizing ecosystem structure. However, in the study of energy flow through the ecosystem, pyramids of energy are the most consistent and representative models of ecosystem structure (**Figure 5.14**).

Ecological Pyramids

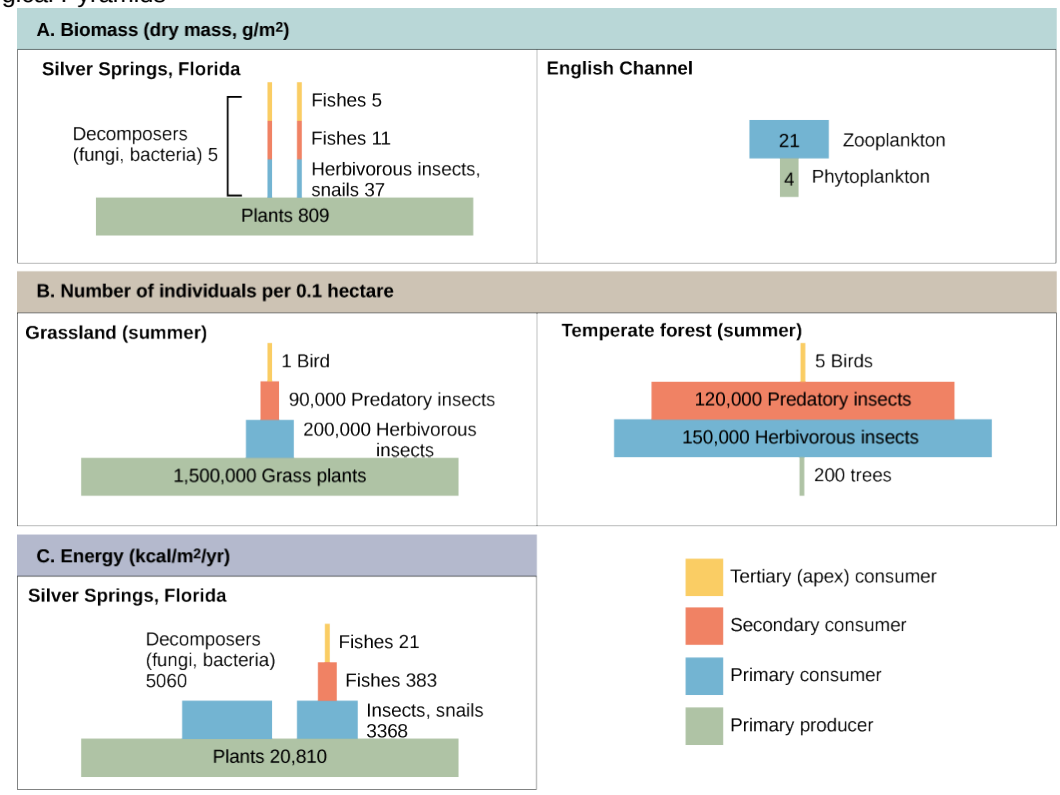


Figure 5.14 Ecological pyramids depict the (a) biomass, (b) number of organisms, and (c) energy in each trophic level.

Consequences of Food Webs: Biological Magnification

One of the most important environmental consequences of ecosystem dynamics is biological magnification. **Biological magnification** is the increasing concentration of persistent, toxic substances in organisms at higher trophic levels, from the primary producers to the apex consumers. Many substances have been shown to bioaccumulate, including the pesticide **d**ichloro**d**iphenyl**t**richloroethane (DDT), which was first publicized in the 1960s bestseller, *Silent Spring*, by Rachel Carson. DDT was a commonly used pesticide before it became known that DDT and its metabolites persist in ecosystems and organisms, and these compounds can have harmful effects on many species in the higher trophic levels. In some aquatic

ecosystems, organisms in higher trophic levels consume many individuals from the trophic level below. Small amounts of toxins in the water become increasingly concentrated (magnified) from lower trophic levels to higher trophic levels. So, DDT occurs in low concentrations in the water is acquired by the producers, then magnified in the primary consumers (small aquatic animals), then magnified again in the secondary consumers (fish), and again in higher consumers such as fish-eating birds (Ospreys, Pelicans, Bald Eagles). High levels of DDT metabolites cause the eggshells of birds to become thin and fragile; they often crack or break long before the baby bird hatches out. This effect resulted in significant declines in populations of fish-eating birds. The use of DDT was banned in the United States in the 1970s.

Other substances that biomagnify are polychlorinated biphenyls (PCBs), which were used in coolant liquids in the United States until their use was banned in 1979, and heavy metals, such as mercury, lead, and cadmium. These substances were best studied in aquatic ecosystems, where fish species at different trophic levels accumulate toxic substances that are found in low concentrations in the primary producers. As illustrated in a study performed by the National Oceanic and Atmospheric Administration (NOAA) in the Saginaw Bay of Lake Huron (**Figure 5.15**), PCB concentrations increased from the ecosystem's primary producers (phytoplankton) through the different trophic levels of fish species. The apex consumer (walleye) has more than four times the amount of PCBs per gram, compared to phytoplankton. Additionally, based on results from other studies, birds that eat these fish may have PCB levels at least one order of magnitude higher than those found in the lake fish.

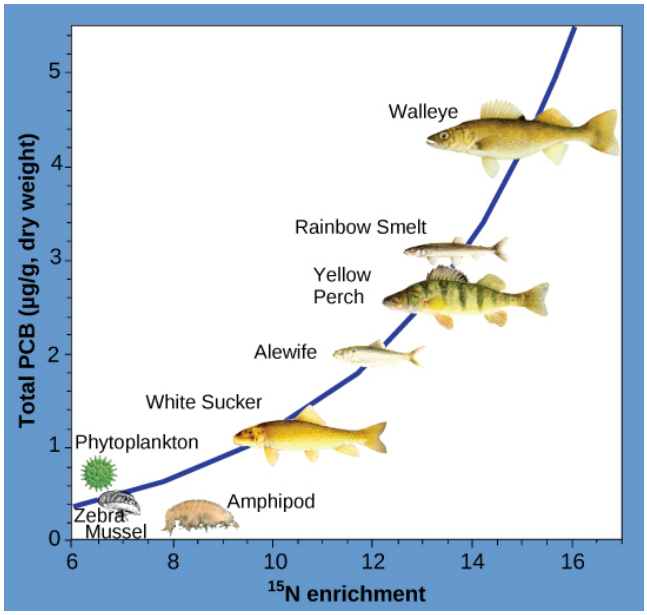


Figure 5.15 This chart shows the PCB concentrations found at the various trophic levels in the Saginaw Bay ecosystem of Lake Huron. Numbers on the x-axis reflect enrichment with heavy isotopes of nitrogen (^{15}N), which is a marker for increasing trophic level. Notice that the fish in the higher trophic levels accumulate more PCBs than those in lower trophic levels. (credit: Patricia Van Hoof, NOAA, GLERL)

Other concerns have been raised by the accumulation of heavy metals, such as mercury and cadmium, in certain types of seafood. The United States Environmental Protection Agency (EPA) recommends that pregnant women and young children should not consume any top predator species such as swordfish, shark, king mackerel, or tilefish, because of their high mercury content. These individuals are advised to eat fish low in mercury: salmon, tilapia, shrimp, pollock, and catfish, which are lower on the trophic pyramids in their ecosystems. Biological magnification is a good example of how ecosystem dynamics can affect our everyday lives, even influencing the food we eat.

5.5 | Biogeochemical Cycles

Energy flows, nutrients cycle

"We are stardust, we are golden, we are billion year old carbon."

- Joni Mitchell

Indeed, the elements that make up our bodies, and those of every other living thing, were born in dying stars, billions of years ago. And energy from our own star, the Sun, is an important player in making those elements available to living organisms on this planet. Energy flows directionally through ecosystems, entering as sunlight (or inorganic molecules for chemoautotrophs) and leaving as heat during the many transfers between trophic levels. However, the matter that makes up living organisms (**nutrients**) is conserved and recycled. The five most common elements associated with organic molecules—carbon, nitrogen, hydrogen, oxygen, and phosphorus—take a variety of chemical forms and may exist for long periods in the atmosphere, on land, in water, or beneath the Earth's surface. Geologic processes, such as weathering, erosion, water drainage, and the subduction of the continental plates, all play a role in this recycling of materials. Because geology and chemistry have major roles in the study of this process, the recycling of inorganic matter between living organisms and their environment is called a **biogeochemical cycle**. It is important to understand that energy, flowing through ecosystems, is also needed to drive biogeochemical cycles.

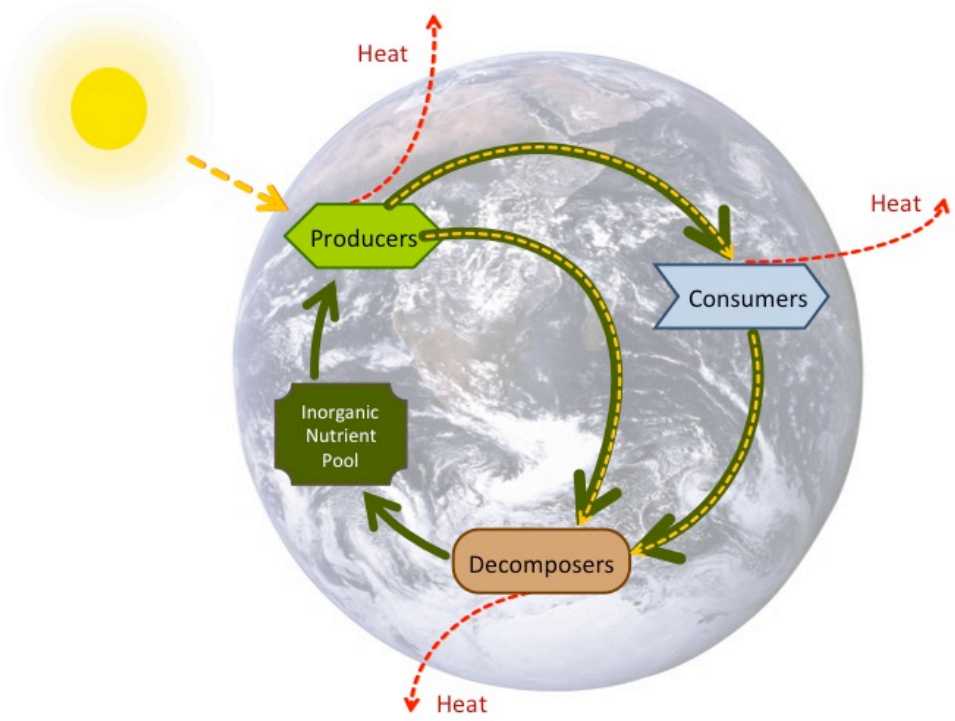


Figure 5.16 This image illustrates the flow of energy and the cycling of nutrients. The dark green lines represent the movement of nutrients and the dashed lines represent the movement of energy. As you can see, nutrients remain within the system while energy enter via photosynthesis and leaves the system primarily as heat energy, a non-biolgically useful form of energy. Work by Eva Horne and Robert A. Bear

Water contains hydrogen and oxygen, which are essential to all living processes. The hydrosphere is the area of the Earth where water movement and storage occurs: as liquid water on the surface and beneath the surface, or frozen (rivers, lakes, oceans, groundwater, polar ice caps, and glaciers), and as water vapor in the atmosphere. Carbon is found in all organic macromolecules and is an important constituent of fossil fuels. Nitrogen is a major component of nucleic acids and proteins and is critical to human agriculture. Phosphorus, a major component of nucleic acids (along with nitrogen), is also one of the main ingredients in artificial fertilizers used in agriculture. Sulfur, an element that is involved in the 3–D folding of proteins (as in disulfide binding), is released into the atmosphere by the burning of fossil fuels, such as coal. Cycling and recycling of these chemicals from the environment to organisms and back again is critically important to all living things.

The cycling of these elements is interconnected. For example, the movement of water is critical for the leaching of nitrogen and phosphate into rivers, lakes, and oceans. Furthermore, the ocean itself is a major reservoir for carbon. Thus, mineral nutrients are cycled, either rapidly or slowly, through the entire biosphere, from one living organism to another, and between the biotic and abiotic world.

The Water (Hydrologic) Cycle

Water is the basis of all living processes. The human body is more than 1/2 water and human cells are more than 70 percent water. Thus, most land animals need a supply of fresh water to survive. However, when examining the stores of water on Earth, 97.5 percent of it is non-potable salt water (Figure 5.17). Of the remaining water, 99 percent is locked underground as water or as ice. Thus, less than 1 percent of fresh water is easily accessible from lakes and rivers. Many living things, such as plants, animals, and fungi, are dependent on the small amount of fresh surface water supply, a lack of which can have massive effects on ecosystem dynamics. Humans, of course, have developed technologies to increase water availability, such as digging wells to harvest groundwater, storing rainwater, and using desalination to obtain drinkable water from the ocean. Although this pursuit of drinkable water has been ongoing throughout human history, the supply of fresh water is still a major issue in modern times.

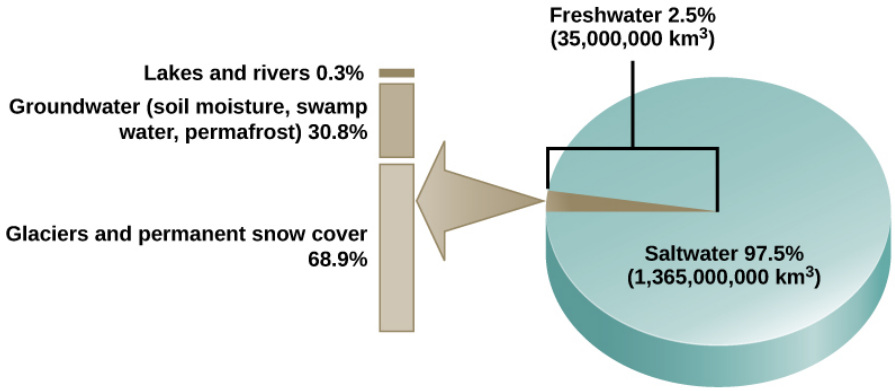


Figure 5.17 Only 2.5 percent of water on Earth is fresh water, and less than 1 percent of fresh water is easily accessible to living things.

Water cycling is extremely important to ecosystem dynamics. Water has a major influence on climate and, thus, on the environments of ecosystems. Most of the water on Earth is stored for long periods in the oceans, underground, and as ice. Figure 5.18 illustrates the average time that an individual water molecule may spend in the Earth's major water reservoirs. Residence time is a measure of the average time an individual water molecule stays in a particular reservoir. A large amount of the Earth's water is locked in place in these reservoirs as ice, beneath the ground, and in the ocean, and, thus, is unavailable for short-term cycling (only surface water can evaporate).

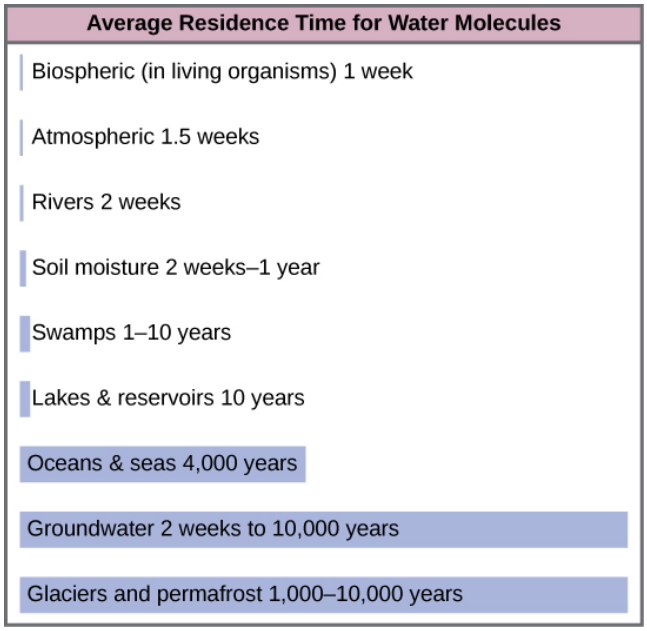


Figure 5.18 This graph shows the average residence time for water molecules in the Earth's water reservoirs.

There are various processes that occur during the cycling of water, shown in Figure 5.19. These processes include the following:

- evaporation/sublimation
- condensation/precipitation
- subsurface water flow
- surface runoff/snowmelt
- streamflow

The water cycle is driven by the sun's energy as it warms the oceans and other surface waters. This leads to the evaporation (water to water vapor) of liquid surface water and the sublimation (ice to water vapor) of frozen water, which deposits large amounts of water vapor into the atmosphere. Over time, this water vapor condenses into clouds as liquid or frozen droplets and is eventually followed by precipitation (rain or snow), which returns water to the Earth's surface. Rain eventually permeates into the ground, where it may evaporate again if it is near the surface, flow beneath the surface, or be stored for long periods. More easily observed is surface runoff: the flow of fresh water either from rain or melting ice. Runoff can then make its way through streams and lakes to the oceans or flow directly to the oceans themselves.

Rain and surface runoff are major ways in which minerals, including carbon, nitrogen, phosphorus, and sulfur, are cycled from land to water. The environmental effects of runoff will be discussed later as these cycles are described.

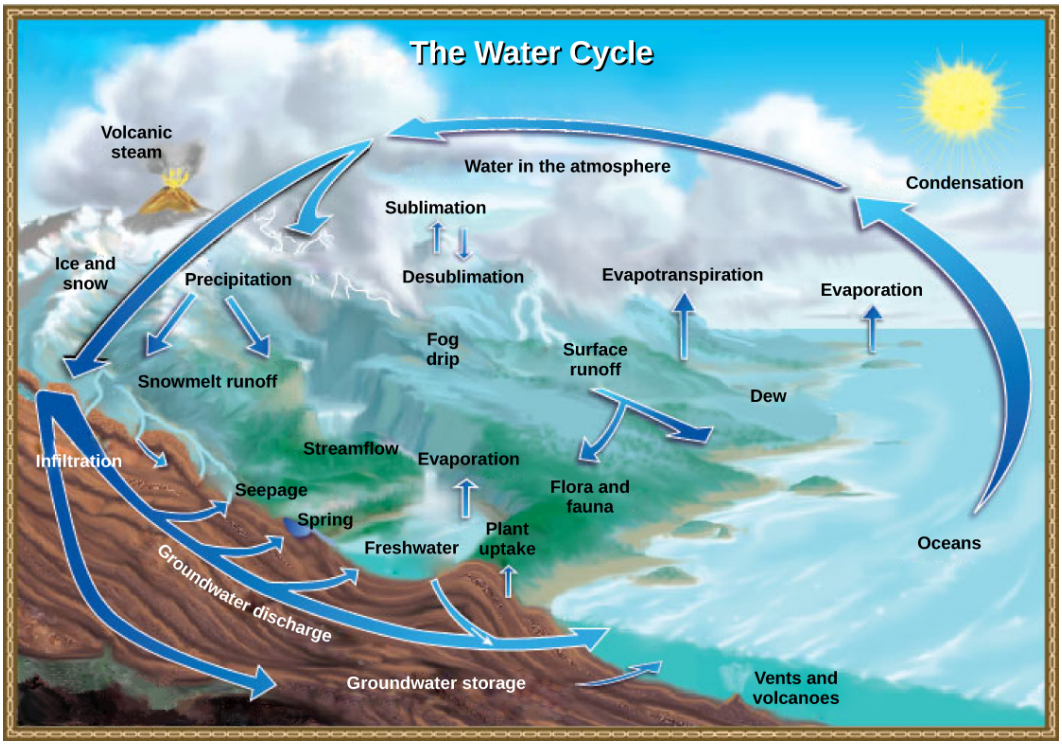


Figure 5.19 Water from the land and oceans enters the atmosphere by evaporation or sublimation, where it condenses into clouds and falls as rain or snow. Precipitated water may enter freshwater bodies or infiltrate the soil. The cycle is complete when surface or groundwater reenters the ocean. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

The Carbon Cycle

Carbon is the second most abundant element in living organisms. Carbon is present in all organic molecules, and its role in the structure of macromolecules is of primary importance to living organisms. Carbon compounds contain especially high energy, particularly those derived from fossilized organisms, mainly plants, which humans use as fuel. Since the 1800s, the number of countries using massive amounts of fossil fuels has increased. Since the beginning of the Industrial Revolution, global demand for the Earth's limited fossil fuel supplies has risen; therefore, the amount of carbon dioxide in our atmosphere has increased. This increase in carbon dioxide has been associated with climate change and other disturbances of the Earth's ecosystems and is a major environmental concern worldwide. Thus, the "carbon footprint" is based on how much carbon dioxide is produced and how much fossil fuel countries consume.

The carbon cycle is most easily studied as two interconnected sub-cycles: one dealing with rapid carbon exchange among living organisms and the other dealing with the long-term cycling of carbon through geologic processes. The entire carbon

cycle is shown in Figure 5.20.

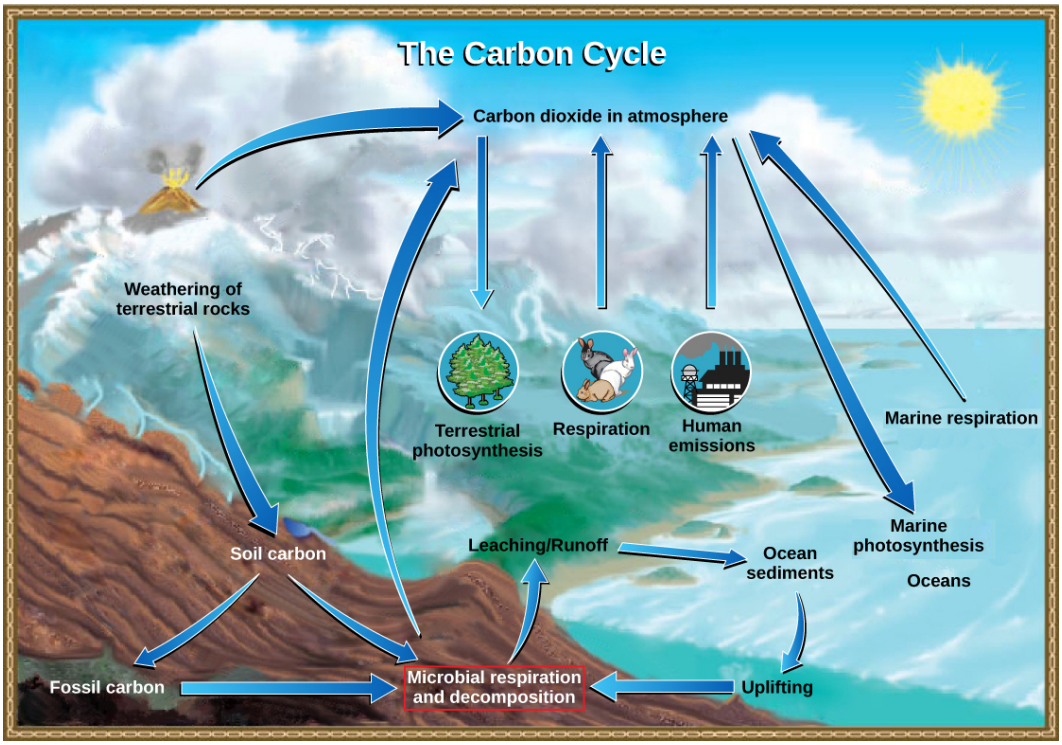


Figure 5.20 Carbon dioxide gas exists in the atmosphere and is dissolved in water. Photosynthesis converts carbon dioxide gas to organic carbon, and respiration cycles the organic carbon back into carbon dioxide gas. Long-term storage of organic carbon occurs when matter from living organisms is buried deep underground and becomes fossilized. Volcanic activity and, more recently, human emissions, bring this stored carbon back into the carbon cycle. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

The Biological Carbon Cycle

Living organisms are connected in many ways, even between ecosystems. A good example of this connection is the exchange of carbon between autotrophs and heterotrophs within and between ecosystems by way of atmospheric carbon dioxide. Carbon dioxide is the basic building block that most autotrophs use to build multi-carbon, high energy compounds, such as glucose. The energy harnessed from the sun is used by these organisms to form the covalent bonds that link carbon atoms together. These chemical bonds thereby store this energy for later use in the process of respiration. Most terrestrial autotrophs obtain their carbon dioxide directly from the atmosphere, while marine autotrophs acquire it in the dissolved form (carbonic acid, $H_2CO_3^-$). However carbon dioxide is acquired, a by-product of the process is oxygen. Photosynthetic organisms are responsible for the approximately 21 percent oxygen content of the atmosphere that we observe today.

Heterotrophs and autotrophs are partners in biological carbon exchange (especially the primary consumers, largely herbivores). Heterotrophs acquire the high-energy carbon compounds from the autotrophs by consuming them, and breaking them down by respiration to obtain cellular energy, such as ATP. The most efficient type of respiration, aerobic respiration, requires oxygen obtained from the atmosphere or dissolved in water. Thus, there is a constant exchange of oxygen and carbon dioxide between the autotrophs (which need the carbon) and the heterotrophs (which need the oxygen). Gas exchange through the atmosphere and water is one way that the carbon cycle connects all living organisms on Earth.

The Biogeochemical Carbon Cycle

The movement of carbon through the land, water, and air is complex, and in many cases, it occurs much more slowly geologically than it does between living organisms. Carbon is stored for long periods in what are known as carbon reservoirs, which include the atmosphere, bodies of liquid water (mostly oceans), ocean sediment, soil, land sediments (including fossil fuels), and the Earth's interior.

As stated, the atmosphere is a major reservoir of carbon in the form of carbon dioxide and is essential to the process of photosynthesis. The level of carbon dioxide in the atmosphere is greatly influenced by the reservoir of carbon in the oceans. The exchange of carbon between the atmosphere and water reservoirs influences how much carbon is found in each location, and each one affects the other reciprocally. Carbon dioxide (CO_2) from the atmosphere dissolves in water and combines with water molecules to form carbonic acid, and then it ionizes to carbonate and bicarbonate ions (Figure 5.21)

Step 1: CO_2 (atmospheric) ⇌ CO_2 (dissolved)
Step 2: CO_2 (dissolved) + H_2O ⇌ H_2CO_3 (carbonic acid)
Step 3: H_2CO_3 ⇌ + H^+ + HCO_3^- (biocarbonate ion)
Step 4: HCO_3^- ⇌ H^+ + CO_3^{2-} (carbonate ion)

Figure 5.21 Carbon dioxide reacts with water to form bicarbonate and carbonate ions.

The equilibrium coefficients are such that more than 90 percent of the carbon in the ocean is found as bicarbonate ions. Some of these ions combine with seawater calcium to form calcium carbonate ($CaCO_3$), a major component of marine organism shells. These organisms eventually form sediments on the ocean floor. Over geologic time, the calcium carbonate forms limestone, which comprises the largest carbon reservoir on Earth.

On land, carbon is stored in soil as a result of the decomposition of living organisms (by decomposers) or from weathering of terrestrial rock and minerals. This carbon can be leached into the water reservoirs by surface runoff. Deeper underground, on land and at sea, are fossil fuels: the anaerobically decomposed remains of plants that take millions of years to form. Fossil fuels are considered a non-renewable resource because their use far exceeds their rate of formation. A **non-renewable resource**, such as fossil fuel, is either regenerated very slowly or not at all. Another way for carbon to enter the atmosphere is from land (including land beneath the surface of the ocean) by the eruption of volcanoes and other geothermal systems. Carbon sediments from the ocean floor are taken deep within the Earth by the process of **subduction**: the movement of one tectonic plate beneath another. Carbon is released as carbon dioxide when a volcano erupts or from volcanic hydrothermal vents.

Carbon dioxide is also added to the atmosphere by the animal husbandry practices of humans. The large numbers of land animals raised to feed the Earth's growing population results in increased carbon dioxide levels in the atmosphere due to farming practices and the respiration and methane production. This is another example of how human activity indirectly affects biogeochemical cycles in a significant way. Although much of the debate about the future effects of increasing atmospheric carbon on climate change focuses on fossils fuels, scientists take natural processes, such as volcanoes and respiration, into account as they model and predict the future impact of this increase.

The Nitrogen Cycle

Nitrogen is an essential nutrient for living processes; it is a major component of proteins and nucleic acids. Proteins are important biological molecules because all cellular activities are driven by proteins. Nucleic Acids are the building blocks of DNA (hereditary material). Nitrogen is often the limiting nutrient (necessary for growth) on terrestrial ecosystems **Figure 5.22**.

Getting nitrogen into the living world is difficult. Plants and phytoplankton are not equipped to incorporate nitrogen from the atmosphere (which exists as tightly bonded, triple covalent N_2) even though this molecule comprises approximately 78 percent of the atmosphere. Nitrogen enters the living world via free-living and symbiotic bacteria, which incorporate nitrogen into their macromolecules through nitrogen fixation (conversion of N_2). Cyanobacteria live in most aquatic ecosystems where sunlight is present; they play a key role in nitrogen fixation. Cyanobacteria are able to use inorganic sources of nitrogen to "fix" nitrogen. *Rhizobium* bacteria live symbiotically in the root nodules of legumes (such as peas, beans, and peanuts) and provide them with the organic nitrogen they need. Free-living bacteria, such as *Azotobacter*, are also important nitrogen fixers. In addition, humans industrially fix nitrogen to produce artificial fertilizers.

Organic nitrogen is especially important to the study of ecosystem dynamics since many ecosystem processes, such as primary production and decomposition, are limited by the available supply of nitrogen. As shown in **Figure 5.22**, the nitrogen that enters living systems by nitrogen fixation is successively converted from organic nitrogen back into nitrogen gas by bacteria. This process occurs in three steps in terrestrial systems: ammonification, nitrification, and denitrification. First, the ammonification process converts nitrogenous waste from living animals or from the remains of dead animals into ammonium (NH_4^+) by certain bacteria and fungi. Second, the ammonium is converted to nitrites (NO_2^-) by nitrifying bacteria, such as *Nitrosomonas*, through nitrification. Subsequently, nitrites are converted to nitrates (NO_3^-) by similar organisms. Third, the process of denitrification occurs, whereby bacteria, such as *Pseudomonas* and *Clostridium*, convert the nitrates into nitrogen gas, allowing it to re-enter the atmosphere.

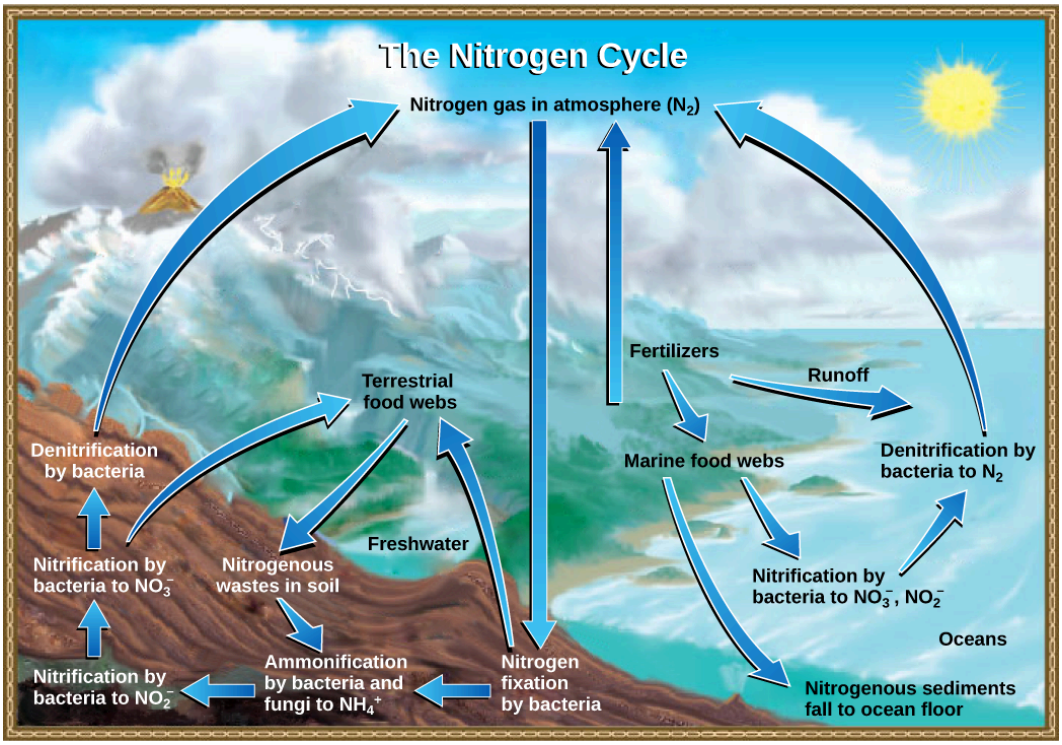


Figure 5.22 Nitrogen enters the living world from the atmosphere via nitrogen-fixing bacteria. This nitrogen and nitrogenous waste from animals is then processed back into gaseous nitrogen by soil bacteria, which also supply terrestrial food webs with the organic nitrogen they need. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

Human activity can release nitrogen into the environment by two primary means: the combustion of fossil fuels, which releases different nitrogen oxides, and by the use of artificial fertilizers in agriculture, which are then washed into lakes, streams, and rivers by surface runoff. Atmospheric nitrogen is associated with several effects on Earth's ecosystems including the production of acid rain (as nitric acid, HNO_3) and greenhouse gas (as nitrous oxide, N_2O) potentially causing climate change. A major effect from fertilizer runoff is saltwater and freshwater **eutrophication**, a process whereby nutrient runoff causes the excess growth of microorganisms, depleting dissolved oxygen levels and killing ecosystem fauna.

A similar process occurs in the marine nitrogen cycle, where the ammonification, nitrification, and denitrification processes are performed by marine bacteria. Some of this nitrogen falls to the ocean floor as sediment, which can then be moved to land in geologic time by uplift of the Earth's surface and thereby incorporated into terrestrial rock. Although the movement of nitrogen from rock directly into living systems has been traditionally seen as insignificant compared with nitrogen fixed from the atmosphere, a recent study showed that this process may indeed be significant and should be included in any study of the global nitrogen cycle.[2]

The Phosphorus Cycle

Phosphorus is another essential nutrient for living processes; it is a major component of nucleic acids, phospholipids, and, as calcium phosphate, makes up the supportive components of our bones. Phosphorus is often the limiting nutrient (necessary for growth) in aquatic ecosystems (**Figure 5.23**).

Phosphorus occurs in nature as the phosphate ion (PO_4^{3-}). In addition to phosphate runoff as a result of human activity (mined and used to make a artificial fertilizer), natural surface runoff occurs when it is leached from phosphate-containing rock by weathering, thus sending phosphates into rivers, lakes, and the ocean. This rock has its origins in the ocean. Phosphate-containing ocean sediments form primarily from the bodies of ocean organisms and from their excretions. However, in remote regions, volcanic ash, aerosols, and mineral dust may also be significant phosphate sources. This sediment then is moved to land over geologic time by the uplifting of areas of the Earth's surface. Phosphorus is .

Phosphorus is also reciprocally exchanged between phosphate dissolved in the ocean and marine ecosystems. The movement of phosphate from the ocean to the land and through the soil is extremely slow, with the average phosphate ion

2. Scott L. Morford, Benjamin Z. Houlton, and Randy A. Dahlgren, "Increased Forest Ecosystem Carbon and Nitrogen Storage from Nitrogen Rich Bedrock," *Nature* 477, no. 7362 (2011): 78–81.

having an oceanic residence time between 20,000 and 100,000 years.

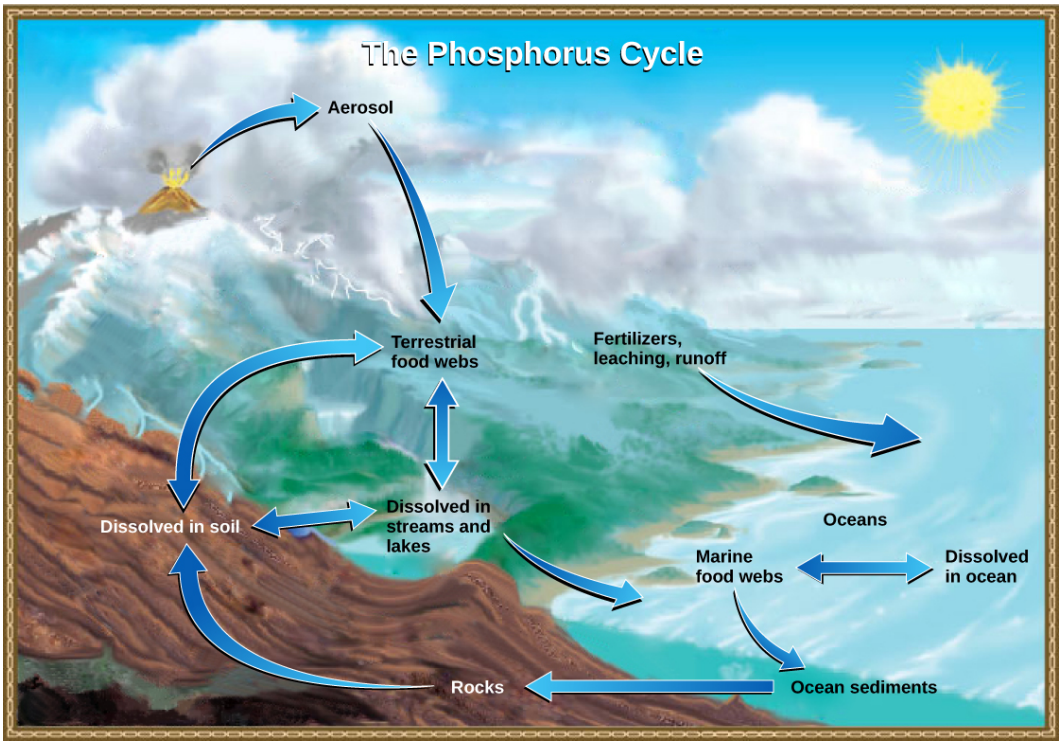


Figure 5.23 In nature, phosphorus exists as the phosphate ion (PO_4^{3-}). Weathering of rocks and volcanic activity releases phosphate into the soil, water, and air, where it becomes available to terrestrial food webs. Phosphate enters the oceans via surface runoff, groundwater flow, and river flow. Phosphate dissolved in ocean water cycles into marine food webs. Some phosphate from the marine food webs falls to the ocean floor, where it forms sediment. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

Excess phosphorus and nitrogen entering these ecosystems from fertilizer runoff, and from sewage, causes excessive growth of microorganisms. This is known as eutrophication (the enrichment of bodies of fresh water by inorganic plant nutrients). These abundant microorganism "blooms" then die and decay, which can increase the rate of sedimentation. But the most important biological effect is to deplete the oxygen which is dissolved in the water. Oxygen is critical for most aquatic organisms, and oxygen depletion leads to the death of many of the larger organisms, such as shellfish and finfish. This process is responsible for dead zones in lakes and at the mouths of many major rivers (**Figure 5.24**).

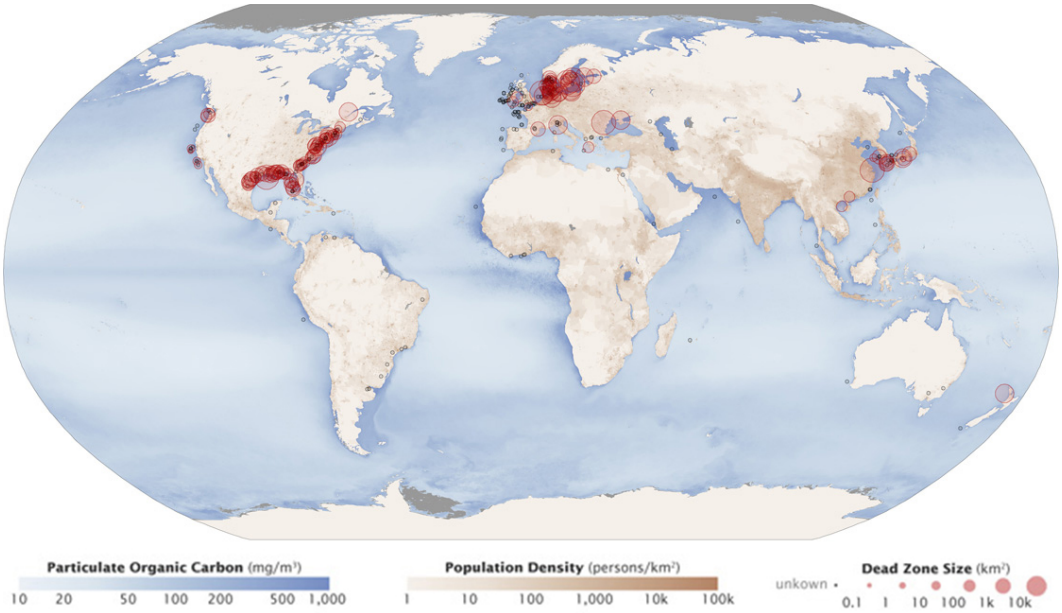


Figure 5.24 Dead zones occur when phosphorus and nitrogen from fertilizers cause excessive growth of microorganisms, which depletes oxygen and kills fauna. Worldwide, large dead zones are found in coastal areas of high population density. (credit: NASA Earth Observatory)

A **dead zone** is an area within a freshwater or marine ecosystem where large areas are depleted of their normal flora and fauna; these zones can be caused by eutrophication, oil spills, dumping of toxic chemicals, and other human activities. The number of dead zones has been increasing for several years, and more than 400 of these zones were present as of 2008. One of the worst dead zones is off the coast of the United States in the Gulf of Mexico, where fertilizer runoff from the Mississippi River basin has created a dead zone of over 8463 square miles. Phosphate and nitrate runoff from fertilizers also negatively affect several lake and bay ecosystems including the Chesapeake Bay in the eastern United States.

5.6 | Biogeography

Introduction

> “Biology is a science of three dimensions. The first is the study of each species across all levels of biological organization, molecule to cell to organism to population to ecosystem. The second dimension is the diversity of all species in the biosphere. The third dimension is the history of each species in turn, comprising both its genetic evolution and the environmental change that drove the evolution. Biology, by growing in all three dimensions, is progressing toward unification and will continue to do so.”

- Edward O. Wilson, in 'Systematics and the Future of Biology', Systematics and the Origin of Species: on Ernst Mayr's 100th anniversary, Volume 102, Issues 22-26 (2005)

Many forces influence the communities of living organisms present in different parts of the biosphere (all of the parts of Earth inhabited by life). The biosphere extends into the atmosphere (several kilometers above Earth) and into the depths of the oceans. Despite its apparent vastness to an individual human, the biosphere occupies only a minute space when compared to the known universe. Many abiotic forces influence where life can exist and the types of organisms found in

different parts of the biosphere. The abiotic factors influence the distribution of **biomes**: large areas of land with similar climate, flora, and fauna.

Biogeography

Biogeography is the study of the geographic distribution of living things and the abiotic factors that affect their distribution. Abiotic factors such as temperature and rainfall vary based mainly on latitude and elevation. As these abiotic factors change, the composition of plant and animal communities also changes. For example, if you were to begin a journey at the equator and walk north, you would notice gradual changes in plant communities. At the beginning of your journey, you would see tropical wet forests with broad-leaved evergreen trees, which are characteristic of plant communities found near the equator. As you continued to travel north, you would see these broad-leaved evergreen plants eventually give rise to seasonally dry forests with scattered trees. You would also begin to notice changes in temperature and moisture. At about 30 degrees north, these forests would give way to deserts, which are characterized by low precipitation.

Moving farther north, you would see that deserts are replaced by grasslands or prairies. Eventually, grasslands are replaced by deciduous temperate forests. These deciduous forests give way to the boreal forests found in the subarctic, the area south of the Arctic Circle. Finally, you would reach the Arctic tundra, which is found at the most northern latitudes. This trek north reveals gradual changes in both climate and the types of organisms that have adapted to environmental factors associated with ecosystems found at different latitudes. However, different ecosystems exist at the same latitude due in part to abiotic factors such as atmospheric jet streams, and other ocean currents. A simplified model of the air circulation at different latitudes, which is a primary cause of the climate and ecozones at different latitudes, is shown in the figure below. There are six rotating cells of air in the upper atmosphere (three north of the equator, and three south of the equator) that, combined with the earth's rotation, create the earth's climate and influence the distribution of plants and animals.

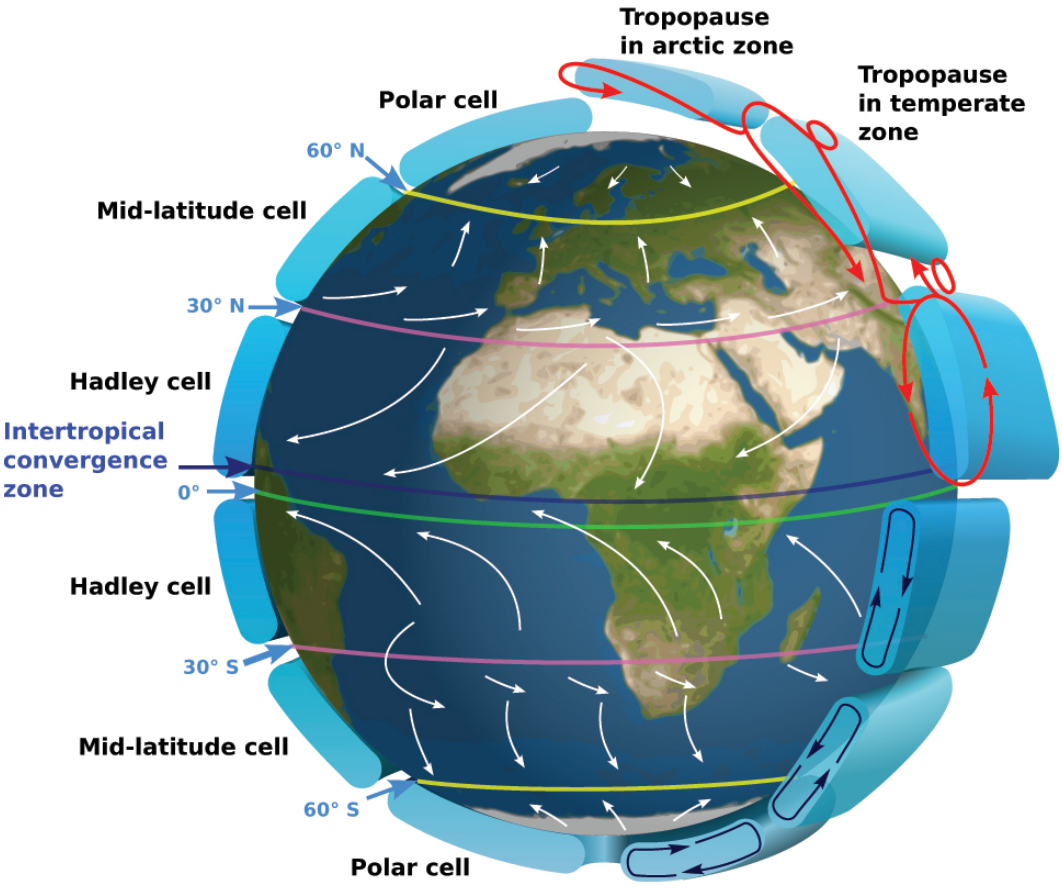


Figure 5.25 Global Air Circulation patterns Simplified diagram of the upper air circulation patterns that help create the earth's climate. (modified from Wikimedia Commons). Rising and descending air creates three "cells" that encircle the globe. High pressure areas are created in regions where the air descends; low pressure areas arise where the air is ascending. Descending air has lost much of its moisture (there are lots of thunderstorms in the tropics!), and so moisture is absorbed in those regions where dryer air returns to the surface (e.g. 30 degrees N and 30 degrees S). This results in a band of desert regions at those latitudes.

Ecologists who study biogeography are especially interested in patterns of species distribution. No species exists

everywhere; most are found in relatively small areas of the world. For example, the Venus flytrap is endemic to a small area in North and South Carolina. An **endemic** species is one which is naturally found only in a specific geographic area that is usually restricted in size. Other species are generalists: species which live in a wide variety of geographic areas; the raccoon, for example, is native to most of North and Central America. Some birds (e.g. Osprey, Pandion halietus) are found in appropriate habitats on several continents.

Species distribution patterns are based on biotic and abiotic factors, and are also influenced by events occurring during the very long periods of time required for species evolution. Early studies of biogeography were closely linked to the emergence of evolutionary thinking in the eighteenth century, and, in fact, observations of these patterns helped Wallace and Darwin formulate the theory of evolution. Some of the most distinctive assemblages of plants and animals occur in regions that have been physically separated for millions of years by geographic barriers. Biologists estimate that Australia, for example, has between 600,000 and 700,000 species of plants and animals; 92% of the plant species, and 83% of the mammal species in Australia are endemic (found on no other continent). See the Figure below for a couple of examples. This is a consequence of the fact that Australia and Asia have been geographically separated for at least 50 million years.

(a) (b)

Figure 5.26 Australia is home to many endemic species. The (a) wallaby (*Wallabia bicolor*), a medium-sized member of the kangaroo family, is a pouched mammal, or marsupial. The (b) echidna (*Tachyglossus aculeatus*) is an egg-laying mammal. (credit a: modification of work by Derrick Coetzee; credit b: modification of work by Allan Whittome)

Sometimes ecologists discover unique patterns of species distribution by determining where species are *not* found. Hawaii, for example, has no native land species of reptiles or amphibians, and has only one native terrestrial mammal, the hoary bat. Most of New Guinea lacks placental mammals, and prior to human settlement, there were no land mammals in New Zealand except for three species of bats.

Plants can be endemic or generalists: endemic plants are found only on specific regions of the Earth, while generalists are found on many regions. Isolated land masses—such as Australia, Hawaii, and Madagascar—often have large numbers of endemic plant species. Some of these plants are endangered due to human activity. The forest gardenia (*Gardenia brighamii*), for instance, is endemic to Hawaii; only an estimated 15–20 trees are thought to exist (**Figure 5.27**).

Figure 5.27 Listed as federally endangered, the forest gardenia is a small tree with distinctive flowers. It is found only in five of the Hawaiian Islands in small populations consisting of a few individual specimens. (credit: Forest & Kim Starr)

Energy Sources

The distribution of organisms is also influenced by the availability of energy in their environment. Energy from the sun is captured by green plants, algae, cyanobacteria, and photosynthetic protists. These organisms convert solar energy into the chemical energy needed by all living things. Light availability can be an important force directly affecting the evolution of adaptations in photosynthesizers. For instance, plants in the understory of a temperate forest are shaded when the trees above them in the canopy completely leaf out in the late spring. Not surprisingly, understory plants have adaptations to successfully capture available light. One such adaptation is the rapid growth of spring ephemeral plants such as the spring beauty (**Figure 5.28**). These spring flowers achieve much of their growth and finish their life cycle (reproduce) early in the season before the trees in the canopy develop leaves.

Figure 5.28 The spring beauty is an ephemeral spring plant that flowers early in the spring to avoid competing with larger forest trees for sunlight. (credit: John Beetham)

In aquatic ecosystems, the availability of light may be limited because sunlight is absorbed by water, plants, suspended particles, and resident microorganisms. Toward the bottom of a lake, pond, or ocean, there may be a zone that light cannot reach. Photosynthesis cannot take place there and, as a result, a number of adaptations have evolved that enable living things to survive in these situations. For instance, aquatic plants have photosynthetic tissue near the surface of the water; the broad, floating leaves of a water lily ensure that this organism gets the light it needs to survive. In totally dark environments such as hydrothermal vents in the deep ocean, some bacteria extract energy from inorganic chemicals using chemosynthesis, a metabolic pathway similar to photosynthesis. You'll learn more about those pathways in a later module.

The availability of inorganic nutrients in aquatic systems is also an important aspect of energy or photosynthesis. Many organisms sink to the bottom of the ocean when they die in the open water; when this occurs, the nutrients and energy in that organism are out of circulation for some time, unless ocean upwelling occurs. **Ocean upwelling** is the rising of deep ocean

waters that occurs when prevailing winds blow along surface waters near a coastline (Figure 5.29). As the wind pushes ocean waters offshore, water from the bottom of the ocean moves up to replace this water. As a result, the nutrients once contained in dead organisms become available for reuse by other living organisms.

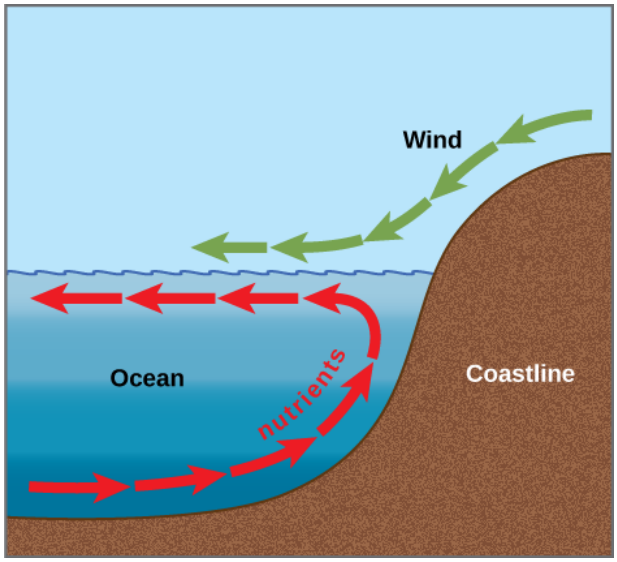


Figure 5.29 Ocean upwelling is an important process that recycles nutrients and energy in the ocean. As wind (green arrows) pushes offshore, it causes water from the ocean bottom (red arrows) to move to the surface, bringing up nutrients from the ocean depths.

In freshwater systems, the recycling of nutrients occurs in response to air temperature changes. The nutrients at the bottom of lakes are recycled twice each year: in the spring and fall turnover. The **spring and fall turnover** is a seasonal process that recycles nutrients and oxygen from the bottom of a freshwater ecosystem to the top (Figure 5.30). These turnovers are caused by the formation of a **thermocline**: a layer of water with a temperature that is significantly different from that of the surrounding layers. In wintertime, the surface of lakes found in many northern regions is frozen. However, the water under the ice is slightly warmer, and the water at the bottom of the lake is warmer yet at 4 °C to 5 °C (39.2 °F to 41 °F). Water is densest at 4 °C; therefore, the deepest water is also the densest. The deepest water is oxygen poor because the decomposition of organic material at the bottom of the lake uses up available oxygen that cannot be replaced in the winter. There is little or no photosynthesis in that season, and any diffusion of oxygen from the atmosphere is blocked by the surface ice layer.

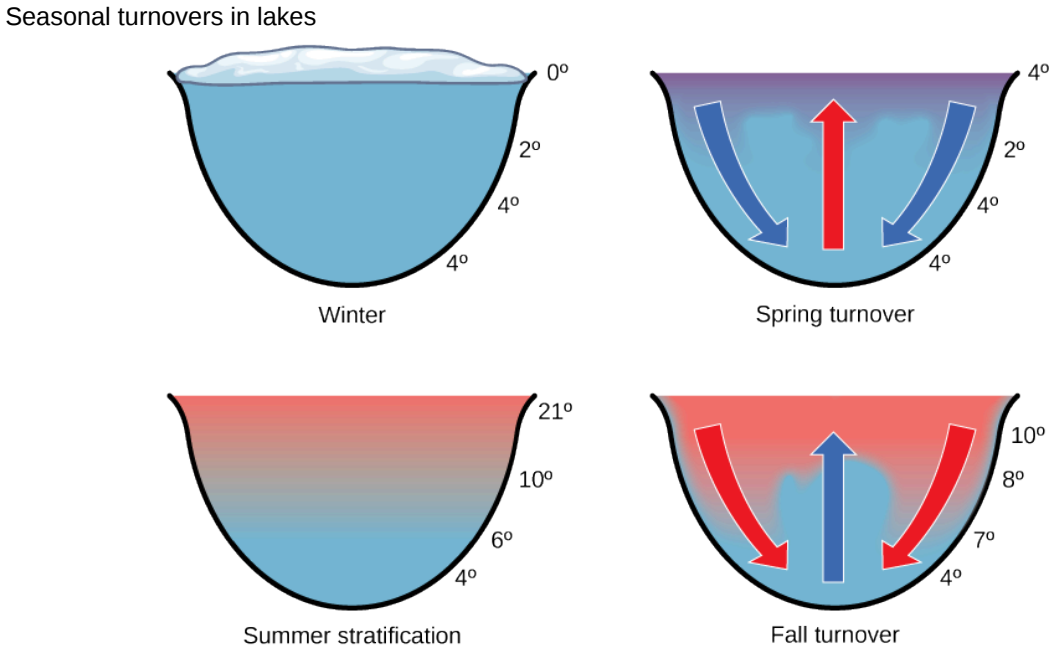


Figure 5.30 The spring and fall turnovers are important processes in freshwater lakes that act to move the nutrients and oxygen at the bottom of deep lakes to the top. Turnover occurs because water has a maximum density at 4 °C. Surface water temperature changes as the seasons progress, and denser water sinks.

In springtime, air temperatures increase and surface ice melts. When the temperature of the surface water begins to reach 4 °C, the water becomes heavier and sinks to the bottom. The water at the bottom of the lake is then displaced by the heavier surface water and, thus, rises to the top. As that water rises to the top, the sediments and nutrients from the lake bottom are brought along with it. During the summer months, the lake water stratifies, or forms layers, with the warmest water at the lake surface.

As air temperatures drop in the fall, the temperature of the lake water cools to 4 °C; therefore, this causes fall turnover as the heavy cold water sinks and displaces the water at the bottom. The oxygen-rich water at the surface of the lake then moves to the bottom of the lake, while the nutrients at the bottom of the lake rise to the surface (**Figure 5.30**). During the winter, the oxygen at the bottom of the lake is used by decomposers and other organisms requiring oxygen, such as fish.

Temperature

In addition to its effects on the density of water, temperature affects the physiology (and the biogeographic distribution) of living things. Temperature exerts an important influence on living things because few organisms can survive at temperatures below 0° C (32° F), due to metabolic constraints. It is also rare for living things to survive at temperatures exceeding 45° C (113° F). This is primarily due to temperature effects on the proteins known as enzymes. Enzymes are typically most efficient within a narrow and specific range of temperatures; enzyme denaturation (damage) can occur at higher temperatures, and enzymes do not work fast enough at lower temperatures. Therefore, organisms either must maintain an internal temperature that keeps their enzymes functioning, or they must inhabit an environment that will keep the body within a temperature range that supports metabolism. Some animals have adapted to enable their bodies to survive significant temperature fluctuations. Some Antarctic fish live at temperatures below freezing, and hibernating Arctic ground squirrels (*Urocitellus parryii*) can survive if their body temperature drops below freezing. Similarly, some bacteria are adapted to surviving in extremely hot environments, such as geysers, boiling mud pits or deep-ocean hydrothermal vents. Such bacteria are examples of extremophiles: organisms that thrive in extreme environments.

Temperature can also limit the distribution of living things. Animals in regions with large temperature fluctuations may respond with various adaptations, such as migration, in order to survive. Migration, the movement from one place to another, is an adaptation found in many animals, including many that inhabit seasonally cold climates. Migration solves problems related to temperature, locating food, and finding a mate. In migration, for instance, the Arctic Tern (*Sterna paradisaea*) makes a 40,000 km (24,000 mi) round trip flight each year between its feeding grounds in the southern hemisphere and its breeding grounds in the Arctic. Monarch butterflies (*Danaus plexippus*) live in the eastern United States in the warmer months and migrate to Mexico and the southern United States in the wintertime. Some species of mammals also make migratory forays. Reindeer (*Rangifer tarandus*) travel about 5,000 km (3,100 mi) each year to find

food. Amphibians and reptiles are more limited in their distribution because they lack migratory ability. Not all animals that can migrate do so: migration carries risk and comes at a high energy cost.

Other successful adaptations allow animals to stay in place and not migrate. Some animals hibernate or estivate to survive hostile temperatures. Hibernation enables animals to survive cold conditions, and estivation allows animals to survive the hostile conditions of a hot, dry climate. Animals that hibernate or estivate enter a state known as torpor: a condition in which their metabolic rate is significantly lowered. This enables the animal to wait until its environment better supports its survival. Some amphibians, such as the wood frog (*Rana sylvatica*), have an antifreeze-like chemical in their cells, which prevents water in the cell from freezing and expanding until the cell is destroyed.

Water

Water is required by all living things because it is critical for cellular processes. Since terrestrial organisms lose water to the environment by simple diffusion, they have evolved many adaptations to retain water.

- Plants have a number of interesting features on their leaves, such as leaf hairs and a waxy cuticle, that serve to decrease the rate of water loss via transpiration.
- Animals that live in very dry environments have many adaptations that allow them to survive without water intake. Kangaroo rats, which live in arid areas in the US and Mexico, obtain most of their water from metabolism of the carbohydrates and lipids in the seeds that make up their diet. They almost never drink water, and would rarely see liquid water in their normal habitats. They also produce very small amounts of highly concentrated, nearly crystalline, urine.
- Freshwater organisms are surrounded by water and are constantly in danger of having water rush into their cells because of osmosis. Many adaptations of organisms living in freshwater environments have evolved to ensure that solute concentrations in their bodies remain within appropriate levels. One such adaptation is the excretion of dilute urine.
- Marine organisms are surrounded by water with a higher solute concentration than the organism and, thus, are in danger of losing water to the environment because of osmosis. These organisms have morphological and physiological adaptations to retain water and release solutes into the environment. For example, Marine Iguanas (*Amblyrhynchus cristatus*), sneeze out water vapor that is high in salt in order to maintain solute concentrations within an acceptable range while swimming in the ocean and eating marine plants.

Inorganic Nutrients and Soil

Inorganic nutrients, such as nitrogen and phosphorus, are important in the distribution and the abundance of living things. Plants obtain these inorganic nutrients from the soil when water moves into the plant through the roots. Therefore, soil structure (particle size of soil components), soil pH, and soil nutrient content play an important role in the distribution of plants. Animals obtain inorganic nutrients from the food they consume. Therefore, animal distributions are related to the distribution of what they eat. In some cases, animals will follow their food resource as it moves through the environment.

Other Aquatic Factors

Some abiotic factors, such as oxygen, are important in aquatic ecosystems as well as terrestrial environments. Terrestrial animals obtain oxygen from the air they breathe. Oxygen availability can be an issue for organisms living at very high elevations, however, where there are fewer molecules of oxygen in the air. In aquatic systems, the concentration of dissolved oxygen is related to water temperature and the speed at which the water moves. Cold water has more dissolved oxygen than warmer water. In addition, salinity, current, and tide can be important abiotic factors in aquatic ecosystems.

Other Terrestrial Factors

Wind can be an important abiotic factor because it influences the rate of evaporation and transpiration. The physical force of wind is also important because it can move soil, water, or other abiotic factors, as well as an ecosystem's organisms.

Fire is another terrestrial factor that can be an important agent of disturbance in terrestrial ecosystems. Some organisms are adapted to fire and, thus, require the high heat associated with fire to complete a part of their life cycle. For example, the jack pine—a coniferous tree—requires heat from fire for its seed cones to open (**Figure 5.31**). Through the burning of pine needles, fire adds nitrogen to the soil and limits competition by destroying undergrowth. Closer to home, the tallgrass prairie ecosystem of the Kansas Flint Hills is dependent on fire and grazing by large herbivores (formerly bison, now cattle). In the absence of such disturbances, the grasslands of the Flint Hills become scrubby cedar forests in just a few decades.

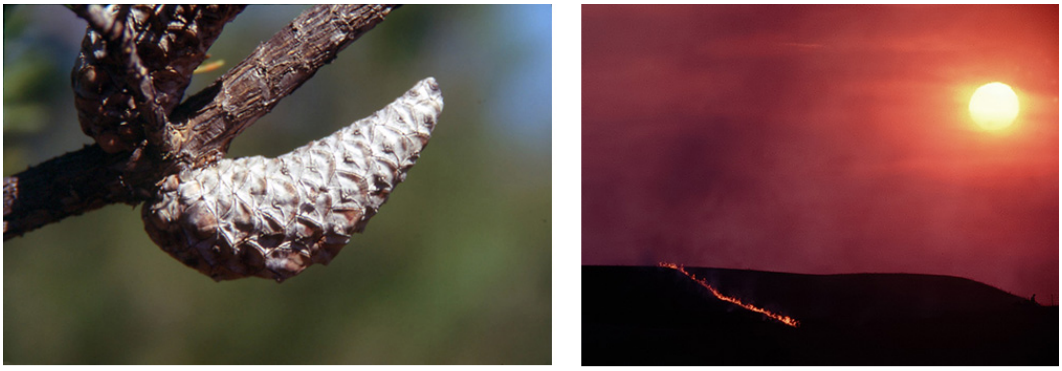


Figure 5.31 (a) The mature cones of the jack pine *(Pinus banksiana)* open only when exposed to high temperatures, such as during a forest fire. A fire is likely to kill most vegetation, so a seedling that germinates after a fire is more likely to receive ample sunlight than one that germinates under normal conditions. (credit: USDA) (b) A controlled burn moves across the Konza Prairie Biological Station. Fire is a critical determinant in the maintenance of the tallgrass prairie ecosystem. (photo by D.A. Rintoul)

Abiotic Factors Influencing Plant Growth

Temperature and moisture are important influences on plant production and the amount of organic matter available to other organisms (net primary productivity). **Net primary productivity** is an estimation of all of the organic matter available to organisms in other trophic levels; it is calculated as the total amount of carbon incorporated into plant tissues per year minus the amount that is used during plant metabolism. In terrestrial environments, net primary productivity is estimated by measuring the aboveground biomass per unit area, which is the total mass of living plants, excluding roots. This means that a large percentage of plant biomass which exists underground is not included in this measurement. Net primary productivity is an important variable when considering differences between biomes. Very productive biomes have a high level of net primary productivity, i.e., a large amount of energy at the primary producer trophic level.

Annual biomass production is directly related to the abiotic components of the environment. Environments with the greatest amount of biomass have conditions in which photosynthesis, plant growth, and the resulting net primary productivity are optimized. The climate of these areas is warm, wet, and usually stable year-round. Photosynthesis can proceed at a high rate, enzymes can work most efficiently, and stomata can remain open without the risk of excessive transpiration. Together, these factors lead to the maximal amount of carbon dioxide (CO_2) moving into the plant, resulting in high biomass production. This biomass produces several important resources for other living things, including habitat and food. Conversely, dry and cold environments have lower photosynthetic rates and therefore less biomass. The animal communities, and the complexity of the food webs, will also be affected by the decrease in available energy at the primary producer level.

5.7 | Biomes

> “The sea, the woods, the mountains, all suffer in comparison with the prairie…The prairie has a stronger hold upon the senses. Its sublimity arises from its unbounded extent, its barren monotony and desolation, its still, unmoved, calm, stern, almost self-confident grandeur, its strange power of deception, its want of echo, and, in fine, its power of throwing a man back upon himself.”

- Albert Pike (1831-32, Journeys in the Prairie)

Figure 5.32 Konza PrairieA grassland Biome - Konza Prairie, near Manhattan, KS. (photograph by David A. Rintoul)

The prairie grassland biome, described by Pike, is one of the Earth's great biomes. But what is a biome, exactly? Biomes are large areas of land (or water) with similar climate, flora, and fauna. The Earth's biomes are categorized into two major groups: terrestrial and aquatic. Terrestrial biomes are based on land, while aquatic biomes include both ocean and freshwater biomes. The eight major terrestrial biomes on Earth are each distinguished by characteristic temperatures and amount of precipitation. Comparing the annual totals of precipitation and fluctuations in precipitation from one biome to another provides clues as to the importance of abiotic factors in the distribution of biomes. Temperature variation on a daily and seasonal basis is also important for predicting the geographic distribution of the biome and the vegetation type in the biome. The distribution of these biomes shows that the same biome can occur in geographically distinct areas with similar climates (**Figure 5.33**).

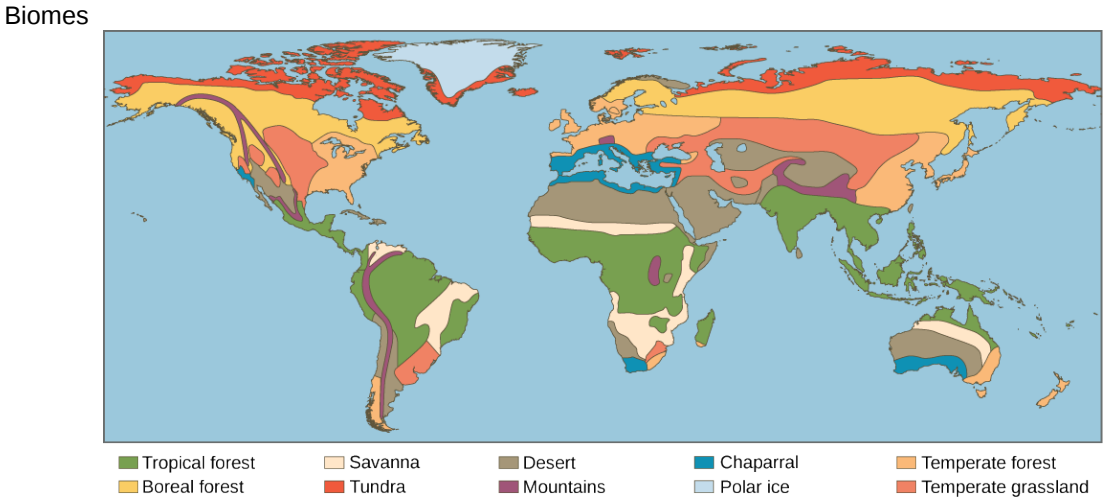


Figure 5.33 Each of the world's major biomes is distinguished by characteristic temperatures and amounts of precipitation. Polar ice and mountains are also shown.

Rainforest

Rainforests are also referred to as tropical rainforests. This biome is found in equatorial regions (**Figure 5.33**). The vegetation is characterized by plants with broad leaves that fall off throughout the year. Unlike the trees of deciduous forests,

the trees in this biome do not have a seasonal loss of leaves associated with variations in temperature and sunlight; these forests are “evergreen” year-round.

The temperature and sunlight profiles of tropical rainforests are very stable in comparison to that of other terrestrial biomes, with the temperatures ranging from 20 °C to 34 °C (68 °F to 93 °F). When one compares the annual temperature variation of tropical rainforests with that of other forest biomes, the lack of seasonal temperature variation in the tropical rainforest becomes apparent. This lack of seasonality leads to year-round plant growth, rather than the seasonal (spring, summer, and fall) growth seen in other biomes. In contrast to other ecosystems, tropical ecosystems do not have long days and short days during the yearly cycle. Instead, a constant daily amount of sunlight (11–12 hrs per day) provides more solar radiation, thereby, a longer period of time for plant growth.

The annual rainfall in tropical rainforests ranges from 125 to 660 cm (50–200 in) with some monthly variation. While sunlight and temperature remain fairly consistent, annual rainfall is highly variable. Tropical rainforests have wet months in which there can be more than 30 cm (11–12 in) of precipitation, as well as dry months in which there are fewer than 10 cm (3.5 in) of rainfall. However, the driest month of a tropical rainforest still exceeds the *annual* rainfall of some other biomes, such as deserts.

Tropical rainforests have high net primary productivity because the annual temperatures and precipitation values in these areas are ideal for plant growth. Therefore, the extensive biomass present in the tropical rainforest leads to plant communities with very high species diversities (**Figure 5.34**). Tropical rainforests have more species of trees than any other biome; on average between 100 and 300 species of trees are present in a single hectare (2.5 acres) of the Amazon region of South America. One way to visualize this is to compare the distinctive horizontal layers within the tropical rainforest biome. On the forest floor is a sparse layer of plants and decaying plant matter. Above that is an understory of short shrubby foliage. A layer of trees rises above this understory and is topped by a closed upper canopy—the uppermost overhead layer of branches and leaves. Some additional trees emerge through this closed upper canopy. These layers provide diverse and complex habitats for the variety of plants, fungi, animals, and other organisms within the tropical rainforests. For instance, epiphytes are plants that grow on other plants, which typically are not harmed. Epiphytes are found throughout tropical rainforest biomes. Many species of animals use the variety of plants and the complex structure of the tropical rainforests for food and shelter. Some organisms live several meters above ground and have adapted to this arboreal lifestyle.

Figure 5.34 Tropical rain forests, such as these forests of Madre de Dios, Peru, near the Amazon River, have high species diversity. (credit: Roosevelt Garcia)

Savannas

Savannas are grasslands with scattered trees, and they are located in Africa, South America, and northern Australia (**Figure 5.33**). Savannas are hot, tropical areas with temperatures averaging from 24 °C to 29 °C (75 °F to 84 °F) and an annual rainfall of 10–40 cm (3.9–15.7 in). Savannas have an extensive dry season; for this reason, forest trees do not grow as well as they do in the tropical wet forest (or other forest biomes). As a result, within the grasses and forbs (herbaceous flowering plants) that dominate the savanna, there are relatively few trees (**Figure 5.35**). Since fire is an important source of disturbance in this biome, plants have evolved well-developed root systems that allow them to quickly re-sprout after a fire.

Figure 5.35 Savannas, like this one in Taita Hills Wildlife Sanctuary in Kenya, are dominated by grasses. (credit: Christopher T. Cooper)

Deserts

Deserts exist between 15 ° and 30 ° north and south latitude and are centered on the Tropics of Cancer and Capricorn (**Figure 5.33**). This biome is very dry; in some years, evaporation exceeds precipitation. Subtropical hot deserts can have daytime soil surface temperatures above 60 °C (140 °F) and nighttime temperatures approaching 0 °C (32 °F). In cold deserts, temperatures can be as high as 25 °C and can drop below -30 °C (-22 °F). Deserts are characterized by low annual precipitation of fewer than 30 cm (12 in) with little monthly variation and lack of predictability in rainfall. In some cases, the annual rainfall can be as low as 2 cm (0.8 in) in deserts located in central Australia (“the Outback”) and northern Africa.

The vegetation and low animal diversity of this biome is closely related to this low and unpredictable precipitation. Very dry deserts lack perennial vegetation that lives from one year to the next; instead, many plants are annuals that grow quickly and reproduce when rainfall does occur, then they die. Many other plants in these areas are characterized by having a number of adaptations that conserve water, such as deep roots, reduced foliage, and water-storing stems (**Figure 5.36**). Seed plants in the desert produce seeds that can be in dormancy for extended periods between rains. Adaptations in desert animals include nocturnal behavior and burrowing.

Figure 5.36 To reduce water loss, many desert plants have tiny leaves or no leaves at all. The leaves of ocotillo (*Fouquieria splendens*), shown here in the Sonora Desert near Gila Bend, Arizona, appear only after rainfall, and then are shed.

Temperate Grasslands

Temperate grasslands are found throughout central North America, where they are also known as prairies; they are also in Eurasia, where they are known as steppes (**Figure 5.33**). Temperate grasslands have pronounced annual fluctuations in temperature with hot summers and cold winters. The annual temperature variation produces specific growing seasons for plants. Plant growth is possible when temperatures are warm enough to sustain plant growth and when ample water is available, which occurs in the spring, summer, and fall. During much of the winter, temperatures are low, and water, which is stored in the form of ice, is not available for plant growth.

Annual precipitation ranges from 25 cm to 75 cm (9.8–29.5 in). Because of relatively lower annual precipitation in temperate grasslands, there are few trees except for those found growing along rivers or streams. The dominant vegetation tends to consist of grasses and some prairies sustain populations of grazing animals **Figure 5.37**. The vegetation is very dense and the soils are fertile because the subsurface of the soil is packed with the roots and rhizomes (underground stems) of these grasses. The roots and rhizomes act to anchor plants into the ground and replenish the organic material (humus) in the soil when they die and decay.

Figure 5.37 The American bison (*Bison bison*), more commonly called the buffalo, is a grazing mammal that once populated American prairies in huge numbers. (photograph by Eva Horne)

Fires, mainly caused by lightning, are a natural disturbance in temperate grasslands. When fire is suppressed in temperate grasslands, the vegetation eventually converts to scrub and dense forests. Often, the restoration or management of temperate grasslands requires the use of controlled burns to suppress the growth of trees and maintain the grasses.

Deciduous Forest

Deciduous forests are the most common biome in eastern North America, Western Europe, Eastern Asia, Chile, and New Zealand (**Figure 5.33**). This biome is found throughout mid-latitude regions. Temperatures range between -30 °C and 30 °C (-22 °F to 86 °F) and drop to below freezing on an annual basis. These temperatures mean that temperate forests have defined growing seasons during the spring, summer, and early fall. Precipitation is relatively constant throughout the year and ranges between 75 cm and 150 cm (29.5–59 in).

Because of the moderate annual rainfall and temperatures, deciduous trees are the dominant plant in this biome (**Figure 5.38**). Deciduous trees lose their leaves each fall and remain leafless in the winter. Thus, no photosynthesis occurs in the deciduous trees during the dormant winter period. Each spring, new leaves appear as the temperature increases. Because of the dormant period, the net primary productivity of temperate forests is less than that of tropical wet forests. In addition, temperate forests show less diversity of tree species than tropical wet forest biomes.

Figure 5.38 Deciduous trees are the dominant plant in the temperate forest. (credit: Oliver Herold)

The trees of the deciduous forests leaf out and shade much of the ground; however, this biome is more open than tropical wet forests because trees in the temperate forests do not grow as tall as the trees in tropical wet forests. The soils of the deciduous forests are rich in inorganic and organic nutrients. This is due to the thick layer of leaf litter on forest floors. As this leaf litter decays, nutrients are returned to the soil. The leaf litter also protects soil from erosion, insulates the ground,

and provides habitats for invertebrates (such as the pill bug or roly-poly, *Armadillidium vulgare*) and their predators, such as the red-backed salamander (*Plethodon cinereus*).

Coniferous Forest

The coniferous forest, also known as taiga or boreal forest, is found south of the Arctic Circle and across most of Canada, Alaska, Russia, and northern Europe (**Figure 5.33**). This biome has cold, dry winters and short, cool, wet summers. The annual precipitation is from 40 cm to 100 cm (15.7–39 in) and usually takes the form of snow. Little evaporation occurs because of the cold temperatures.

The long and cold winters in the coniferous forest have led to the predominance of cold-tolerant cone-bearing plants. These are evergreen coniferous trees like pines, spruce, and fir, which retain their needle-shaped leaves year-round. Evergreen trees can photosynthesize earlier in the spring than deciduous trees because less energy from the sun is required to warm a needle-like leaf than a broad leaf. This benefits evergreen trees, which grow faster than deciduous trees in the coniferous forest. In addition, soils in coniferous forest regions tend to be acidic with little available nitrogen. Leaves are a nitrogen-rich structure and deciduous trees must produce a new set of these nitrogen-rich structures each year. Therefore, coniferous trees that retain nitrogen-rich needles may have a competitive advantage over the broad-leafed deciduous trees.

The net primary productivity of coniferous forests is lower than that of deciduous forests and tropical rain forests. The above ground biomass of coniferous forests is high because these slow-growing tree species are long lived and accumulate standing biomass over time. Plant species diversity is less than that seen in deciduous forests and tropical rain forests. Coniferous forests lack the pronounced elements of the layered forest structure seen in tropical wet forests. The structure of a coniferous forest is often only a tree layer and a ground layer (**Figure 5.39**). When conifer needles are dropped, they decompose more slowly than broad leaves; therefore, fewer nutrients are returned to the soil to fuel plant growth.

Figure 5.39 The coniferous forest (taiga) has low lying plants and conifer trees. (credit: L.B. Brubaker)

Arctic Tundra

The Arctic tundra lies north of the subarctic boreal forest and is located throughout the Arctic regions of the northern hemisphere (**Figure 5.33**). The average winter temperature is -34 °C (-34 °F) and the average summer temperature is from 3 °C to 12 °C (37 °F–52 °F). Plants in the arctic tundra have a very short growing season of approximately 10–12 weeks. However, during this time, there are almost 24 hours of daylight and plant growth is rapid. The annual precipitation of the Arctic tundra is very low with little annual variation in precipitation. And, as in the boreal forests, there is little evaporation due to the cold temperatures.

Plants in the Arctic tundra are generally low to the ground (**Figure 5.40**). There is little species diversity, low net primary productivity, and low aboveground biomass. Deeper soils of the Arctic tundra may remain in a perennially frozen state referred to as permafrost. The permafrost makes it impossible for roots to penetrate deep into the soil and slows the decay of organic matter, which inhibits the release of nutrients from organic matter. During the growing season, the ground of the Arctic tundra can be completely covered with plants or lichens.

Figure 5.40 Low-growing plants such as shrub willow dominate the tundra landscape, shown here in the Arctic National Wildlife Refuge. (credit: USFWS Arctic National Wildlife Refuge)

5.8 | Aquatic Biomes

Abiotic Factors Influencing Aquatic Biomes

Introduction

> "There's nothing wrong with enjoying looking at the surface of the ocean itself, except that when you finally see what goes on underwater, you realize that you've been missing the whole point of the ocean. Staying on the surface all the time is like going to the circus and staring at the outside of the tent."

– Dave Barry

Like terrestrial biomes, aquatic biomes are influenced by a series of abiotic factors. The aquatic medium—water— has different physical and chemical properties than air, however. Even if the water in a pond or other body of water is perfectly clear (there are no suspended particles), water, on its own, absorbs light. As one descends into a deep body of water, there will eventually be a depth which the sunlight cannot reach. While there are some abiotic and biotic factors in a terrestrial ecosystem that might obscure light (like fog, dust, or insect swarms), usually these are not permanent features of the environment. The importance of light in aquatic biomes is central to the communities of organisms found in both freshwater and marine ecosystems. In freshwater systems, stratification due to differences in density is perhaps the most critical abiotic factor and is related to the energy aspects of light. The thermal properties of water (rates of heating and cooling) are significant to the function of marine systems and have major impacts on global climate and weather patterns. Marine systems are also influenced by large-scale physical water movements, such as currents; these are less important in most freshwater lakes.

The ocean is categorized by several areas or zones (**Figure 5.41**). All of the ocean's open water is referred to as the pelagic realm. The benthic realm extends along the ocean bottom from the shoreline to the deepest parts of the ocean floor. Within the pelagic realm is the photic zone, which is the portion of the ocean that light can penetrate (approximately 200 m or 650 ft). At depths greater than 200 m, light cannot penetrate; thus, this is referred to as the aphotic zone. The majority of the ocean is aphotic and lacks sufficient light for photosynthesis. The deepest part of the ocean, the Challenger Deep (in the Mariana Trench, located in the western Pacific Ocean), is about 11,000 m (about 6.8 mi) deep. To give some perspective on the depth of this trench, the ocean is, on average, 4267 m or 14,000 ft deep. These realms and zones are relevant to freshwater lakes as well.

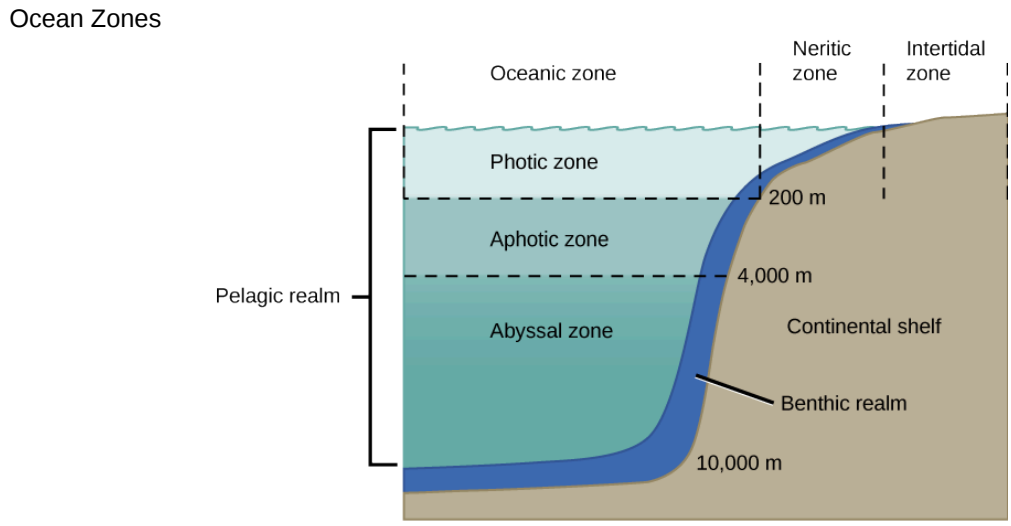


Figure 5.41 The ocean is divided into different zones based on water depth and distance from the shoreline.

Marine Biomes

The ocean is the largest marine biome. It is a continuous body of salt water that is relatively uniform in chemical composition; it is a weak solution of mineral salts and decayed biological matter. Within the ocean, coral reefs are a second kind of marine biome. Estuaries, coastal areas where salt water and fresh water mix, form a third unique marine biome.

Ocean

The physical diversity of the ocean is a significant influence on plants, animals, and other organisms. The ocean is categorized into different zones based on how far light reaches into the water. Each zone has a distinct group of species adapted to the biotic and abiotic conditions particular to that zone.

The intertidal zone, which is the zone between high and low tide, is the oceanic region that is closest to land (**Figure 5.41**). Generally, most people think of this portion of the ocean as a sandy beach. In some cases, the intertidal zone is indeed a sandy beach, but it can also be rocky or muddy. The intertidal zone is an extremely variable environment because of tides. Organisms are exposed to air and sunlight at low tide and are underwater during high tide. Therefore, living things that thrive in the intertidal zone are adapted to being dry for long periods of time. The shore of the intertidal zone is also repeatedly struck by waves, and the organisms found there are adapted to withstand damage from the pounding action of the waves (**Figure 5.42**). The exoskeletons of shoreline crustaceans (such as the shore crab, *Carcinus maenas*) are tough and protect them from desiccation (drying out) and wave damage. Another consequence of the pounding waves is that few algae and plants establish themselves in the constantly moving rocks, sand, or mud.

Figure 5.42 Sea urchins, mussel shells, and starfish are often found in the intertidal zone, shown here in Kachemak Bay, Alaska. (credit: NOAA)

The neritic zone (**Figure 5.41**) extends from the intertidal zone to depths of about 200 m (or 650 ft) at the edge of the continental shelf. Since light can penetrate this depth, photosynthesis can occur in the neritic zone. The water here contains

silt and is well-oxygenated, low in pressure, and stable in temperature. Phytoplankton and floating *Sargassum* (a type of free-floating marine seaweed) provide a habitat for some sea life found in the neritic zone. Zooplankton, protists, small fishes, and shrimp are found in the neritic zone and are the base of the food chain for most of the world's fisheries.

Beyond the neritic zone is the open ocean area known as the oceanic zone (**Figure 5.41**). Within the oceanic zone there is thermal stratification where warm and cold waters mix because of ocean currents. Abundant plankton serve as the base of the food chain for larger animals such as whales and dolphins. Nutrients are scarce and this is a relatively less productive part of the marine biome. When photosynthetic organisms and the protists and animals that feed on them die, their bodies fall to the bottom of the ocean where they remain; unlike freshwater lakes, the open ocean lacks a process for bringing the organic nutrients back up to the surface. The majority of organisms in the aphotic zone include sea cucumbers (phylum Echinodermata) and other organisms that survive on the nutrients contained in the dead bodies of organisms in the photic zone.

Beneath the pelagic zone is the benthic realm, the deepwater region beyond the continental shelf (**Figure 5.41**). The bottom of the benthic realm is comprised of sand, silt, and dead organisms. Temperature decreases, remaining above freezing, as water depth increases. This is a nutrient-rich portion of the ocean because of the dead organisms that fall from the upper layers of the ocean. Because of this high level of nutrients, a diversity of fungi, sponges, sea anemones, marine worms, sea stars, fishes, and bacteria exist.

The deepest part of the ocean is the abyssal zone, which is at depths of 4000 m or greater. The abyssal zone (**Figure 5.41**) is very cold and has very high pressure, high oxygen content, and low nutrient content. There are a variety of invertebrates and fishes found in this zone, but the abyssal zone does not have plants because of the lack of light. Hydrothermal vents are found primarily in the abyssal zone; chemosynthetic bacteria utilize the hydrogen sulfide and other minerals emitted from the vents as an energy source and serve as the base of the food chain found in the abyssal zone.

Coral Reefs

Coral reefs are ocean ridges formed by marine invertebrates living in warm shallow waters within the photic zone of the ocean. They are found within 30˚ north and south of the equator. The Great Barrier Reef is a well-known reef system located several miles off the northeastern coast of Australia. Other coral reef systems are fringing islands, which are directly adjacent to land, or atolls, which are circular reef systems surrounding a former landmass that is now underwater. The coral organisms (members of phylum Cnidaria) are colonies of saltwater polyps that secrete a calcium carbonate skeleton. These calcium-rich skeletons slowly accumulate, forming the underwater reef (**Figure 5.43**). Corals found in shallower waters (at a depth of approximately 60 m or about 200 ft) have a mutualistic relationship with photosynthetic unicellular algae. The relationship provides corals with the majority of the nutrition and the energy they require. The waters in which these corals live are nutritionally poor and, without this mutualism, it would not be possible for large corals to grow. Some corals living in deeper and colder water do not have a mutualistic relationship with algae; these corals attain energy and nutrients using stinging cells on their tentacles to capture prey.

It is estimated that more than 4,000 fish species inhabit coral reefs. These fishes can feed on coral, the cryptofauna (invertebrates found within the calcium carbonate substrate of the coral reefs), or the seaweed and algae that are associated with the coral. In addition, some fish species inhabit the boundaries of a coral reef; these species include predators, herbivores, or planktivores. Predators are animal species that hunt and are carnivores or "flesh eaters." Herbivores eat plant material, and planktivores eat plankton.

Figure 5.43 Coral reefs are formed by the calcium carbonate skeletons of coral organisms, which are marine invertebrates in the phylum Cnidaria. (credit: Terry Hughes)

evolution CONNECTION

Global Decline of Coral Reefs

It takes a long time to build a coral reef. The animals that create coral reefs have evolved over millions of years, continuing to slowly deposit the calcium carbonate that forms their characteristic ocean homes. Bathed in warm tropical waters, the coral animals and their symbiotic algal partners evolved to survive at the upper limit of ocean water temperature.

Together, climate change and human activity pose dual threats to the long-term survival of the world's coral reefs. As global warming due to fossil fuel emissions raises ocean temperatures, coral reefs are suffering. The excessive warmth causes the reefs to expel their symbiotic, food-producing algae, resulting in a phenomenon known as bleaching. When bleaching occurs, the reefs lose much of their characteristic color as the algae and the coral animals die if loss of the symbiotic zooxanthellae is prolonged.

Rising levels of atmospheric carbon dioxide further threaten the corals in other ways; as CO_2 dissolves in ocean waters, it lowers the pH and increases ocean acidity. As acidity increases, it interferes with the calcification that normally occurs as coral animals build their calcium carbonate homes.

When a coral reef begins to die, species diversity plummets as animals lose food and shelter. Coral reefs are also economically important tourist destinations, so the decline of coral reefs poses a serious threat to coastal economies.

Human population growth has damaged corals in other ways, too. As human coastal populations increase, the runoff of sediment and agricultural chemicals has increased, too, causing some of the once-clear tropical waters to become cloudy. At the same time, overfishing of popular fish species has allowed the predator species that eat corals to go unchecked.

Although a rise in global temperatures of 1–2°C (a conservative scientific projection) in the coming decades may not seem large, it is very significant to this biome. When change occurs rapidly, species can become extinct before evolution leads to new adaptations. Many scientists believe that global warming, with its rapid (in terms of evolutionary time) and inexorable increases in temperature, is tipping the balance beyond the point at which many of the world's coral reefs can recover.

Estuaries: Where the Ocean Meets Fresh Water

Estuaries are biomes that occur where a source of fresh water, such as a river, meets the ocean. Therefore, both fresh water and salt water are found in the same vicinity; mixing results in a diluted (brackish) saltwater. Estuaries form protected areas where many of the young offspring of crustaceans, mollusks, and fish begin their lives. Salinity is a very important factor that influences the organisms and the adaptations of the organisms found in estuaries. The salinity of estuaries varies and is based on the rate of flow of its freshwater sources. Once or twice a day, high tides bring salt water into the estuary. Low tides occurring at the same frequency reverse the current of salt water.

The short-term and rapid variation in salinity due to the mixing of fresh water and salt water is a difficult physiological challenge for the plants and animals that inhabit estuaries. Many estuarine plant species are halophytes: plants that can tolerate salty conditions. Halophytic plants are adapted to deal with the salinity resulting from saltwater on their roots or from sea spray. In some halophytes, filters in the roots remove the salt from the water that the plant absorbs. Animals, such as mussels and clams (phylum Mollusca), have developed behavioral adaptations that expend a lot of energy to function in this rapidly changing environment. When these animals are exposed to low salinity, they stop feeding, close their shells, and switch from aerobic respiration (in which they use gills) to anaerobic respiration (a process that does not require oxygen). When high tide returns to the estuary, the salinity and oxygen content of the water increases, and these animals open their shells, begin feeding, and return to aerobic respiration.

Freshwater Biomes

Freshwater biomes include lakes and ponds (standing water) as well as rivers and streams (flowing water). They also include wetlands, which will be discussed later. Humans rely on freshwater biomes to provide aquatic resources for drinking water, crop irrigation, sanitation, and industry. These various roles and human benefits are referred to as ecosystem services. Lakes and ponds are found in terrestrial landscapes and are, therefore, connected with abiotic and biotic factors influencing these terrestrial biomes.

Lakes and Ponds

Lakes and ponds can range in area from a few square meters to thousands of square kilometers. Temperature is an important abiotic factor affecting living things found in lakes and ponds. In the summer, thermal stratification of lakes and ponds occurs when the upper layer of water is warmed by the sun and does not mix with deeper, cooler water. Light can penetrate within the photic zone of the lake or pond. Phytoplankton (algae and cyanobacteria) are found here and carry out photosynthesis, providing the base of the food web of lakes and ponds. Zooplankton, such as rotifers and small crustaceans, consume these phytoplankton. At the bottom of lakes and ponds, bacteria in the aphotic zone break down dead organisms that sink to the bottom.

Nitrogen and phosphorus are important limiting nutrients in lakes and ponds. Because of this, they are determining factors in the amount of phytoplankton growth in lakes and ponds. When there is a large input of nitrogen and phosphorus (from sewage and runoff from fertilized lawns and farms, for example), the growth of algae skyrockets, resulting in a large accumulation of algae called an **algal bloom**. Algal blooms (**Figure 5.44**) can become so extensive that they reduce light penetration in water. As a result, the lake or pond becomes aphotic and photosynthetic plants cannot survive. When the algae die and decompose, severe oxygen depletion of the water occurs. Fishes and other organisms that require oxygen are then more likely to die, and resulting dead zones are found across the globe. Lake Erie and the Gulf of Mexico represent freshwater and marine habitats where phosphorus control and storm water runoff pose significant environmental challenges.

Figure 5.44 The uncontrolled growth of algae in this lake has resulted in an algal bloom. (credit: Jeremy Nettleton)

Rivers and Streams

Rivers and streams are continuously moving bodies of water that carry large amounts of water from the source, or headwater, to a lake or ocean. The largest rivers include the Nile River in Africa, the Amazon River in South America, and the Mississippi River in North America.

Abiotic features of rivers and streams vary along the length of the river or stream. Streams begin at a point of origin referred to as source water. The source water is usually cold, low in nutrients, and clear. The channel (the width of the river or stream) is narrower than at any other place along the length of the river or stream. Because of this, the current is often faster here than at any other point of the river or stream.

The fast-moving water results in minimal silt accumulation at the bottom of the river or stream; therefore, the water is clear. Photosynthesis here is mostly attributed to algae that are growing on rocks; the swift current inhibits the growth of phytoplankton. An additional input of energy can come from leaves or other organic material that falls into the river or stream from trees and other plants that border the water. When the leaves decompose, the organic material and nutrients in the leaves are returned to the water. Plants and animals have adapted to this fast-moving water. For instance, leeches (phylum Annelida) have elongated bodies and suckers on both ends. These suckers attach to the substrate, keeping the leech anchored in place. Freshwater trout species (phylum Chordata) are an important predator in these fast-moving rivers and streams.

As the river or stream flows away from the source, the width of the channel gradually widens and the current slows. This slow-moving water, caused by the gradient decrease and the volume increase as tributaries unite, has more sedimentation. Phytoplankton can also be suspended in slow-moving water. Therefore, the water will not be as clear as it is near the source. The water is also warmer. Worms (phylum Annelida) and insects (phylum Arthropoda) can be found burrowing into the mud. The higher order predator vertebrates (phylum Chordata) include waterfowl, frogs, and fishes. These predators must find food in these slow moving, sometimes murky, waters and, unlike the trout in the waters at the source, these vertebrates may not be able to use vision as their primary sense to find food. Instead, they are more likely to use taste or chemical cues to find prey.

Wetlands

Wetlands are environments in which the soil is either permanently or periodically saturated with water. Wetlands are different from lakes because wetlands are shallow bodies of water whereas lakes vary in depth. Emergent vegetation consists of wetland plants that are rooted in the soil but have portions of leaves, stems, and flowers extending above the water's surface. There are several types of wetlands including marshes, swamps, bogs, mudflats, and salt marshes (**Figure 5.45**). The three shared characteristics among these types—what makes them wetlands—are their hydrology, hydrophytic vegetation, and hydric soils.

Figure 5.45 Located in southern Florida, Everglades National Park is vast array of wetland environments, including sawgrass marshes, cypress swamps, and estuarine mangrove forests. Here, a great egret walks among cypress trees. (credit: NPS)

Freshwater marshes and swamps are characterized by slow and steady water flow. Bogs develop in depressions where water flow is low or nonexistent. Bogs usually occur in areas where there is a clay bottom with poor percolation. Percolation is the movement of water through the pores in the soil or rocks. The water found in a bog is stagnant and oxygen depleted because the oxygen that is used during the decomposition of organic matter is not replaced. As the oxygen in the water is depleted, decomposition slows. This leads to organic acids and other acids building up and lowering the pH of the water. At a lower pH, nitrogen becomes unavailable to plants. This creates a challenge for plants because nitrogen is an important limiting resource. Some types of bog plants (such as sundews, pitcher plants, and Venus flytraps) capture insects and extract the nitrogen from their bodies. Bogs have low net primary productivity because the water found in bogs has low levels of nitrogen and oxygen.

6 | POPULATION ECOLOGY

6.1 | Population

> “Anyone who believes in indefinite growth of anything physical on a physically finite planet is either a madman or an economist.”

Kenneth Boulding, economist (President Kennedy's Environmental Advisor 1966)

In biology, a **population** is a very specific thing. A population is all the members of a species living within a specific area. Populations are typically dynamic entities. They expand and contract, but, as noted above, they cannot expand infinitely. Populations fluctuate based on a number of factors: seasonal and yearly changes in the environment, natural disasters such as forest fires and volcanic eruptions, competition for resources between and within species, and the amount of **habitat** (where an organism lives). The statistical study of population dynamics, **demography**, uses a series of mathematical tools to investigate how populations respond to changes in their biotic and abiotic environments. Many of these tools were originally designed to study human populations. For example, **life tables**, which detail the life expectancy of individuals within a population, were initially developed by life insurance companies to set insurance rates. In fact, while the term “demographics” is commonly used when discussing humans, all living populations can be studied using this approach.

Population Size and Density

The study of any population usually begins by determining how many individuals of a particular species exist, and how closely associated they are with each other. Within a particular habitat, a population can be characterized by its **population size (*N*)**, the total number of individuals, and its **population density**, the number of individuals within a specific area or volume. Population size and density are the two main characteristics used to describe and understand populations. For example, populations with more individuals may be more stable than smaller populations based on their genetic variability, and thus their potential to adapt to the environment. Alternatively, a member of a population with low population density (more spread out in the habitat), might have more difficulty finding a mate to reproduce compared to a population of higher density. As is shown in **Figure 6.1**, smaller organisms tend to be more densely distributed than larger organisms.

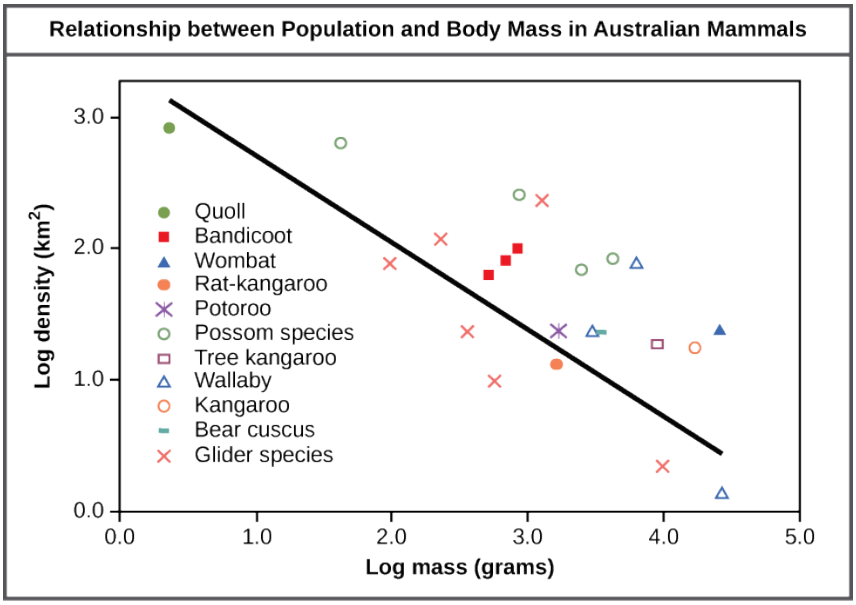


Figure 6.1 Australian mammals show a typical inverse relationship between population density and body size.

Population Research Methods

The most accurate way to determine population size is to simply count all of the individuals within the habitat. However, this method is often not logistically or economically feasible, especially when studying large habitats. Thus, scientists usually study populations by sampling a representative portion of each habitat and using these data to make inferences about the habitat as a whole. A variety of methods can be used to sample populations to determine their size and density. For immobile organisms such as plants, or for very small and slow-moving organisms, a quadrat may be used (**Figure 6.2**). A quadrat is a way of marking off square areas within a habitat, either by staking out an area with sticks and string, or by the use of a wood, plastic, or metal square placed on the ground. After setting the quadrats, researchers then count the number of individuals that lie within their boundaries. Multiple quadrat samples are performed throughout the habitat at several random locations. All of these data can then be used to estimate the population size and population density within the entire habitat. The number and size of quadrat samples depends on the type of organisms under study and other factors, including the density of the organism. For example, if sampling daffodils, a 1 m^2 quadrat might be used whereas with giant redwoods, which are larger and live much further apart from each other, a larger quadrat of 400 m^2 might be employed. This ensures that enough individuals of the species are counted to get an accurate sample that correlates with the habitat, including areas not sampled.

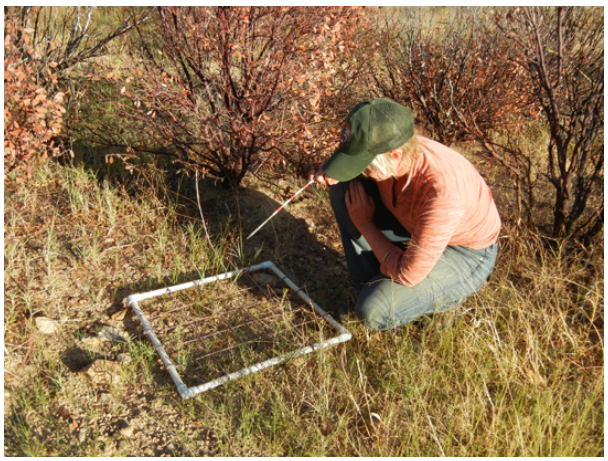
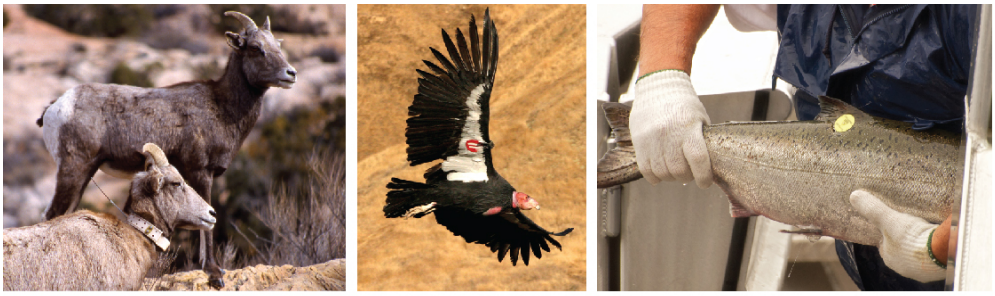


Figure 6.2 A scientist uses a quadrat to measure population size and density. (credit: NPS Sonoran Desert Network)

For mobile organisms, such as mammals, birds, or fish, a technique called **mark and recapture** is often used. This method involves marking a sample of captured animals in some way (such as tags, bands, paint, or other body markings), and then releasing them back into the environment to allow them to mix with the rest of the population; later, a new sample is

collected, including some individuals that are marked (recaptures) and some individuals that are unmarked (Figure 6.3).



(a) (b) (c)

Figure 6.3 Mark and recapture is used to measure the population size of mobile animals such as (a) bighorn sheep, (b) the California condor, and (c) salmon. (credit a: modification of work by Neal Herbert, NPS; credit b: modification of work by Pacific Southwest Region USFWS; credit c: modification of work by Ingrid Taylar)

Using the ratio of marked and unmarked individuals, scientists determine how many individuals are in the sample. From this, calculations are used to estimate the total population size. This method assumes that the larger the population, the lower the percentage of tagged organisms that will be recaptured since they will have mixed with more untagged individuals. For example, if 80 deer are captured, tagged, and released into the forest, and later 100 deer are captured and 20 of them are already marked, we can determine the population size (N) using the following equation:

$$\frac{\text{(number marked first catch x total number of second catch)}}{\text{number marked second catch}} = N$$

Using our example, the population size would be estimated at 400.

$$\frac{(80 \text{ x } 100)}{20} = 400$$

Therefore, there are an estimated 400 total individuals in the original population.

There are some limitations to the mark and recapture method. Some animals from the first catch may learn to avoid capture in the second round, thus inflating population estimates. Alternatively, animals may preferentially be retrapped (especially if a food reward is offered), resulting in an underestimate of population size. Also, some species may be harmed by the marking technique, reducing their survival. A variety of other techniques have been developed, including the electronic tracking of animals tagged with radio transmitters and the use of data from commercial fishing and trapping operations to estimate the size and health of populations and communities.

Species Distribution

In addition to measuring simple density, further information about a population can be obtained by looking at the distribution of the individuals. **Species dispersion patterns** (or distribution patterns) show the spatial relationship between members of a population within a habitat at a particular point in time. In other words, they show whether members of the species live close together or far apart, and what patterns are evident when they are spaced apart.

Individuals in a population can be more or less equally spaced apart, dispersed randomly with no predictable pattern, or clustered in groups. These are known as uniform, random, and clumped dispersion patterns, respectively (Figure 6.4). Uniform dispersion is observed in plants that secrete substances inhibiting the growth of nearby individuals (such as the release of toxic chemicals by the sage plant *Salvia leucophylla*, a phenomenon called allelopathy or plants in arid habitats that compete intensely for limited soil water and nutrients.) and in animals like the penguin that maintain a defined territory. An example of random dispersion occurs with dandelion and other plants that have wind-dispersed seeds that germinate wherever they happen to fall in a favorable environment. A clumped dispersion may be seen in plants that drop their seeds straight to the ground, such as oak trees, or animals that live in groups (schools of fish or herds of elephants). Clumped dispersions may also be a function of habitat heterogeneity. Thus, the dispersion of the individuals within a population provides more information about how they interact with each other than does a simple density measurement. Just as lower density species might have more difficulty finding a mate, solitary species with a random distribution might have a similar difficulty when compared to social species clumped together in groups.

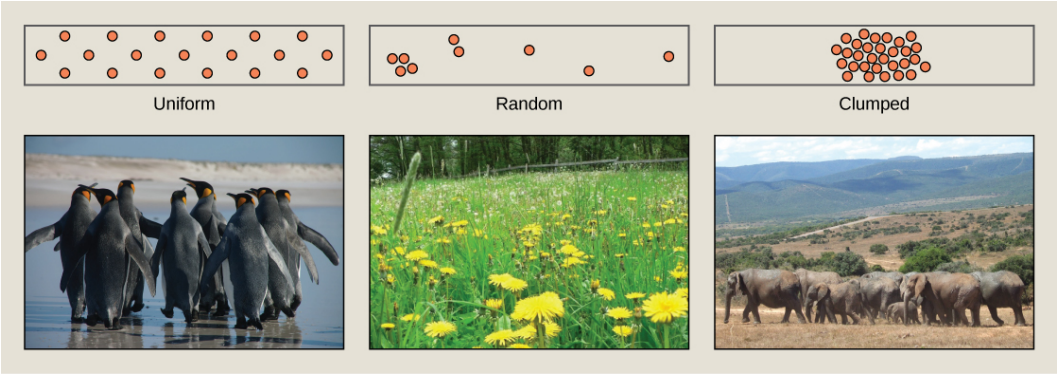


Figure 6.4 Species may have uniform, random, or clumped distribution. Territorial birds such as penguins tend to have uniform distribution. Plants such as dandelions with wind-dispersed seeds tend to be randomly distributed. Animals such as elephants that travel in groups exhibit clumped distribution. (credit a: modification of work by Ben Tubby; credit b: modification of work by Rosendahl; credit c: modification of work by Rebecca Wood)

Demography

While population size and density describe a population at one particular point in time, scientists must use demography to study the dynamics of a population. Demography is the statistical study of population changes over time: birth rates, death rates, migration and life expectancies. Each of these measures, especially birth rates, may be affected by the population characteristics described above. For example, a large population size results in a higher birth rate because more potentially reproductive individuals are present. In contrast, a large population size can also result in a higher death rate because of competition, disease, and the accumulation of waste. Similarly, a higher population density or a clumped dispersion pattern results in more potential reproductive encounters between individuals, which can increase birth rate. Migration the movement of individuals into a population (immigration) and out of a population (emigration) can drastically influence population growth, especially the colonization of new areas. Lastly, a female-biased sex ratio (the ratio of males to females) or age structure (the proportion of population members at specific age ranges) composed of many individuals of reproductive age can increase birth rates.

In addition, the demographic characteristics of a population can influence how the population grows or declines over time. If birth and death rates are equal, the population remains stable. However, the population size will increase if birth rates exceed death rates; the population will decrease if birth rates are less than death rates. Life expectancy is another important factor; the length of time individuals remain in the population impacts local resources, reproduction, and the overall health of the population. These demographic characteristics are often displayed in the form of a life table.

Life Tables

Life tables provide important information about the life history of an organism. Life tables divide the population into age groups and often sexes, and show how long a member of that group is likely to live. They are modeled after actuarial tables used by the insurance industry for estimating human life expectancy. Life tables may include the probability of individuals dying before their next birthday (i.e., their mortality rate, the percentage of surviving individuals dying at a particular age interval, and their life expectancy at each interval. An example of a life table is shown in **Table 6.1** from a study of Dall mountain sheep, a species native to northwestern North America. Notice that the population is divided into age intervals (column A). The mortality rate (per 1000), shown in column D, is based on the number of individuals dying during the age interval (column B) divided by the number of individuals surviving at the beginning of the interval (Column C), multiplied by 1000.

$$\text{mortality rate} = \frac{\text{number of individuals dying}}{\text{number of individuals surviving}} \text{ x } 1000$$

For example, between ages three and four, 12 individuals die out of the 776 that were remaining from the original 1000 sheep. This number is then multiplied by 1000 to get the mortality rate per thousand.

$$\text{mortality rate} = \frac{12}{776} \text{ x } 1000 \approx 15.5$$

As can be seen from the mortality rate data (column D), a high death rate occurred when the sheep were between 6 and 12 months old, and then increased even more from 8 to 12 years old, after which there were few survivors. The data indicate that if a sheep in this population were to survive to age one, it could be expected to live another 7.7 years on average, as shown by the life expectancy numbers in column E.

Life Table of Dall Mountain Sheep[1]

Age interval (years)	Number dying in age interval out of 1000 born	Number surviving at beginning of age interval out of 1000 born	Mortality rate per 1000 alive at beginning of age interval	Life expectancy or mean lifetime remaining to those attaining age interval
0-0.5	54	1000	54.0	7.06
0.5-1	145	946	153.3	--
1-2	12	801	15.0	7.7
2-3	13	789	16.5	6.8
3-4	12	776	15.5	5.9
4-5	30	764	39.3	5.0
5-6	46	734	62.7	4.2
6-7	48	688	69.8	3.4
7-8	69	640	107.8	2.6
8-9	132	571	231.2	1.9
9-10	187	439	426.0	1.3
10-11	156	252	619.0	0.9
11-12	90	96	937.5	0.6
12-13	3	6	500.0	1.2
13-14	3	3	1000	0.7

Table 6.1 This life table of *Ovis dalli* shows the number of deaths, number of survivors, mortality rate, and life expectancy at each age interval for the Dall mountain sheep.

Survivorship Curves

Another tool used by population ecologists is a **survivorship curve**, which is a graph of the number of individuals surviving at each age interval plotted versus time (usually with data compiled from a life table). These curves allow us to compare the life histories of different populations (**Figure 6.5**). Humans and most primates exhibit a Type I survivorship curve because a high percentage of offspring survive their early and middle years—death occurs predominantly in older individuals. These types of species usually have small numbers of offspring at one time, and they give a high amount of parental care to them to ensure their survival. Birds and perennial herbaceous plants are examples of an intermediate of Type II survivorship curve because the probability of death is equal for all age classes in these organisms. Some of the animals also may have relatively few offspring and provide significant parental care. Trees, marine invertebrates, and most fishes exhibit a Type III survivorship curve because very few of these organisms survive their younger years; however, those that make it to an old age are more likely to survive for a relatively long period of time. Organisms in this category usually have a very large number of offspring, but once they are born, little parental care is provided. Thus these offspring are “on their own” and vulnerable to predation, but their sheer numbers assure the survival of enough individuals to perpetuate the species.

1. Data Adapted from Edward S. Deevey, Jr., “Life Tables for Natural Populations of Animals,” *The Quarterly Review of Biology* 22, no. 4 (December 1947): 283-314.

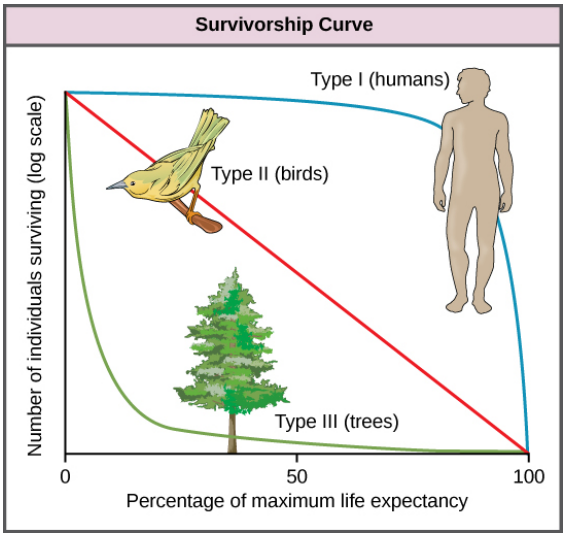


Figure 6.5 Survivorship curves show the distribution of individuals in a population according to age. Humans and most mammals have a Type I survivorship curve because death primarily occurs in the older years. Birds have a Type II survivorship curve, as death at any age is equally probable. Trees have a Type III survivorship curve because very few survive the younger years, but after a certain age, individuals are much more likely to survive.

6.2 | Population Growth

> "The greatest shortcoming of the human race is our inability to understand the exponential function."

- Dr. Albert A. Bartlett, Emeritus Professor of Physics, University of Colorado

Exponential Growth

Charles Darwin, in his theory of evolution by natural selection, was greatly influenced by the English clergyman Thomas Malthus. Malthus published a book in 1798 stating that populations with unlimited natural resources grow very rapidly. This accelerating pattern of increasing population size is called **exponential growth**.

The best example of exponential growth is seen in bacteria. Bacteria can undergo cell division about every hour. If 1000 bacteria are placed in a large flask with an unlimited supply of nutrients (so the nutrients will not become depleted), after an hour, there is one round of division, resulting in 2000 organisms—an increase of 1000. In another hour, each of the 2000 organisms will double, producing 4000, an increase of 2000 organisms. After the third hour, there should be 8000 bacteria in the flask, an increase of 4000 organisms. The important concept of exponential growth is that **population growth (G)**—the number of organisms added in each reproductive generation—is accelerating; that is, it is increasing at a greater and greater rate. After 1 day and 24 of these cycles, the population would have increased from 1000 to more than 16 billion. When the population size, *N*, is plotted over time, a **J-shaped growth curve** is produced (**Figure 6.6**).

The bacteria example is not representative of the real world where resources are limited. Furthermore, some bacteria will die during the experiment and thus not reproduce, lowering the growth rate. Therefore, when calculating the growth of a population, the number of deaths (*D* (number organisms that die during a particular time interval) is subtracted from the number of births (*B*) (number organisms that are born during that interval). This is shown in the following formula:

$$G(\text{population growth}) = B\ (\text{births}) - D\ (\text{deaths})$$

Now let's examine how the average number of births and deaths relate to population growth. The average birth and death rates based on the number of individuals in a population is on a per capita basis. So, the per capita birth rate (*b*) is the number of births during a time interval divided by the number of individuals in the population at that time, and the per capita death rate (*d*) is the number of deaths during a time interval divided by the number of individuals in the population at that time. See the equations below for a simple formula.

$$b(\text{per capita birth rate}) = \frac{B(\text{number of births})}{\text{N (total number of individuals)}}$$

$$d(\text{per capita death rate}) = \frac{D(\text{number of deaths})}{\text{N (total number of individuals)}}$$

Now returning to the simple population growth equation above, we can convert this simple model to one in which births and deaths are expressed on a per capita basis for a time interval. Thus, *B* (births) = *bN* (the per capita birth rate "*b*" multiplied by the number of individuals "*N*") and *D* (deaths) =*dN* (the per capita death rate "*d*" multiplied by the number of individuals "*N*"). When substituting *bN* for *B* and *dN* for *D* in simple growth equation, we can examine the population growth rate based on per capita birth and death rates as seen the equation below.

$$G = bN - dN$$

The above equation can be simplified to the following.

$$G=(b - d)\ \text{N}$$

Now, the difference between per capita birth rate and per capita death rates is further simplified by substituting the term "*r*" (per capita growth rate). So, the per capita growth rate (*r*) is equal to the per capita birth rate (*b*) minus the per capita death rate (*d*), or *r* = *b* - *d*.

$$G = rN$$

The value "*r*" can be positive, meaning the population is increasing in size; or negative, meaning the population is decreasing in size; or zero, where the population's size is unchanging, a condition known as zero population growth. A further refinement of the formula recognizes that different species have inherent differences in their per capita growth rate (often thought of as the potential for reproduction), even under ideal conditions. Obviously, a bacterium can reproduce more rapidly and have a higher per capita growth rate than a human. The maximal growth rate for a species is its biotic potential.

Other factors that influence population growth (G)

When calculating per capita growth rate (r) as above, we are referring to a closed system where individuals from other populations are not moving in and individuals are not leaving. Immigration and emigration are two more factors that can influence the growth of a population. **Immigration** is the movement of individuals into a population which adds individuals to the population, and **emigration** is the movement of individuals out of a population which removes individuals to a population. So, the rates of immigration and emigration are two other factors population biologists need to consider when describing population growth (G).

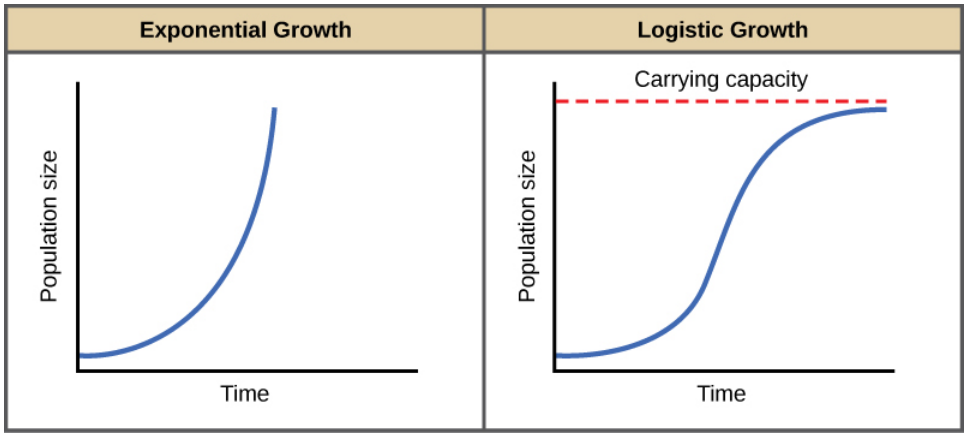


Figure 6.6 When resources are unlimited, populations exhibit exponential growth, resulting in a J-shaped curve. When resources are limited, populations exhibit logistic growth. In logistic growth, population expansion decreases as resources become scarce, and it levels off when the carrying capacity of the environment is reached, resulting in an S-shaped curve.

Logistic Growth

Exponential growth is possible only when natural resources are not limited. This occurs only infrequently and briefly in nature, such as when a population colonizes a new habitat or is recovering from a major disturbance. Charles Darwin recognized this fact in his description of the "struggle for existence," which states that individuals will compete (with members of their own or other species) for limited resources. The successful ones will survive to pass on their own

characteristics and traits (which we know now are transferred by genes) to the next generation at a greater rate (natural selection). To model the reality of limited resources, population ecologists developed the **logistic growth** model.

Carrying Capacity and the Logistic Model

In the real world, with its limited resources, exponential growth cannot continue indefinitely. Exponential growth may occur in environments where there are few individuals and plentiful resources, but when the number of individuals gets large enough, resources will be depleted, slowing the growth rate. Eventually, the growth rate will plateau or level off (**Figure 6.6**). This population size, which represents the maximum population size that a particular environment can support, is called the **carrying capacity**, or *K*.

The formula we use to calculate logistic growth adds the carrying capacity as a moderating force in the growth rate. The expression "*K* – *N*" divided by "*K*" is the fraction of the carrying capacity available for further growth. Thus, the exponential growth model is restricted by this factor to generate the logistic growth equation:

$$G = rN\frac{(K - N)}{K}$$

Notice that when *N* is very small, (*K*-*N*)/*K* becomes close to *K*/*K* or 1, and the right side of the equation reduces to *rN*, which means the population is growing exponentially and is not influenced by carrying capacity. On the other hand, when *N* is large, (*K*-*N*)/*K* comes close to zero, which means that population growth will be slowed greatly or even stopped. Thus, population growth is greatly slowed when the population size is close to the carrying capacity. This model also allows for negative population growth, or a population decline. This occurs when the number of individuals in the population exceeds the carrying capacity (because the value of (K-N)/K is negative).

A graph of this equation yields an **S-shaped curve** (**Figure 6.6**), and it is a more realistic model of population growth than exponential growth. There are three different sections to an S-shaped curve. Initially, growth is exponential because there are few individuals and ample resources available. Then, as resources begin to become limited, the growth rate decreases. Finally, growth levels off at the carrying capacity of the environment, with little change in population size over time.

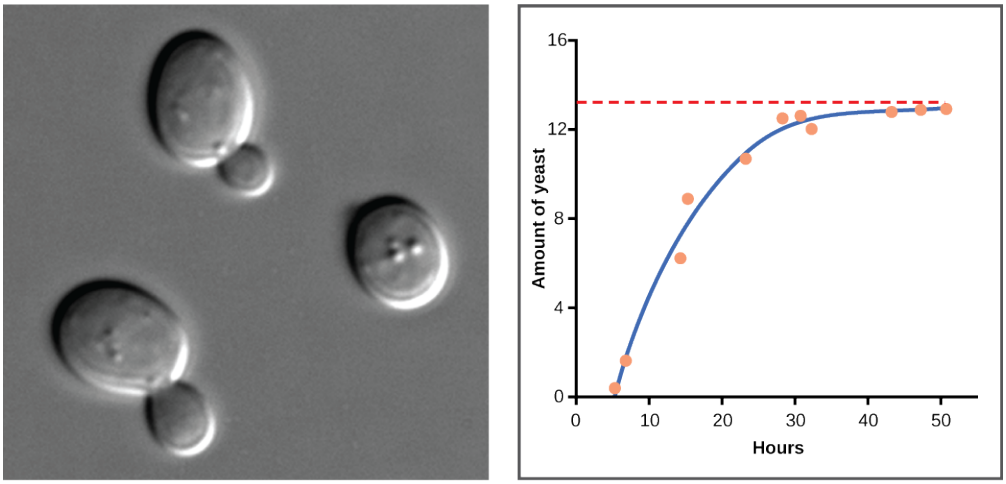
Role of Intraspecific Competition

The logistic model assumes that every individual within a population will have equal access to resources and, thus, an equal chance for survival. For plants, the amount of water, sunlight, nutrients, and the space to grow are the important resources, whereas in animals, important resources may include food, water, shelter, nesting space, and mates.

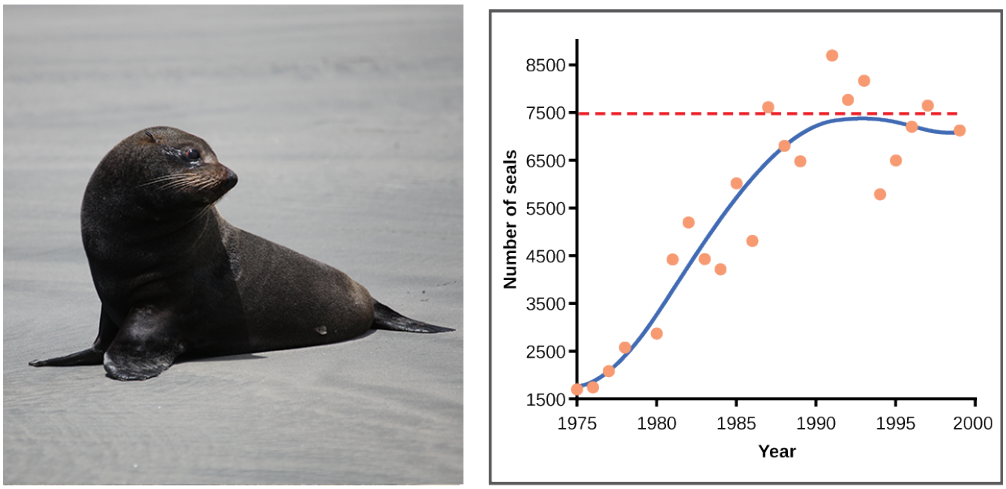
In the real world, variation in a trait among individuals within a population means that some individuals will be better adapted to their environment than others. The resulting competition between population members of the same species for resources is termed **intraspecific competition** (intra- = "within"; -specific = "species"). Intraspecific competition for resources may not affect populations that are well below their carrying capacity—resources are plentiful and all individuals can obtain what they need. However, as population size increases, this competition intensifies. In addition, the accumulation of waste products can reduce an environment's carrying capacity.

Examples of Logistic Growth

Yeast, a microscopic fungus used to make bread and alcoholic beverages, exhibits the classical S-shaped curve when grown in a test tube (**Figure 6.7a**). Its growth levels off as the population depletes the nutrients that are necessary for its growth. However, there are variations to this idealized curve. Examples in wild populations include sheep and harbor seals (**Figure 6.7b**). In both examples, the population size exceeds the carrying capacity for short periods of time and then falls below the carrying capacity afterwards. This fluctuation in population size continues to occur as the population oscillates around its carrying capacity. Still, even with this oscillation, the logistic model is still widely supported.



(a)



(b)

Figure 6.7 (a) Yeast grown in ideal conditions in a test tube show a classical S-shaped logistic growth curve, whereas (b) a natural population of seals shows real-world fluctuation.

6.3 | Population Regulation

> "If we don't halt population growth with justice and compassion, it will be done for us by nature, brutally and without pity- and will leave a ravaged world."

- Dr. Henry W. Kendall, Nobel Laureate in Physics, 1990

The natural mechanisms regulating population growth can indeed be brutal. Carrying capacity is a mathematical sanitization of the realities of famine, disease, and other causes of death. And the logistic model of population growth, while valid in many natural populations and a useful model, is a simplification of real-world population dynamics. Implicit in the model is that the carrying capacity of the environment does not change, which is not the case. The carrying capacity varies annually: for example, some summers are hot and dry whereas others are cold and wet. In many areas, the carrying capacity during the winter is much lower than it is during the summer. Also, natural events such as earthquakes, volcanoes, and fires can alter

an environment and hence its carrying capacity. Additionally, populations do not usually exist in isolation. They engage in **interspecific competition**: that is, they share the environment with other species, competing with them for the same resources. These factors are also important to understanding how a specific population will grow.

Nature regulates population growth in a variety of ways. These are grouped into **density-dependent** factors, in which the density of the population at a given time affects growth rate and mortality, and **density-independent** factors, which influence mortality in a population regardless of population density. Note that in the former, the effect of the factor on the population depends on the density of the population at onset. Conservation biologists want to understand both types because this helps them manage populations and prevent extinction or overpopulation.

Density-dependent Regulation

Most density-dependent factors are biological in nature (biotic), and include predation, inter- and intraspecific competition, accumulation of waste, and diseases such as those caused by parasites. Usually, the denser a population is, the greater its mortality rate. For example, during intra- and interspecific competition, the reproductive rates of the individuals will usually be lower and mortality rates higher, reducing their population's rate of growth. In addition, low prey density increases the mortality of its predator because it has more difficulty locating its food source.

An example of density-dependent regulation is shown in **Figure 6.8** with results from a study focusing on the giant intestinal roundworm (*Ascaris lumbricoides*), a parasite of humans and other mammals.[2] Denser populations of the parasite exhibited lower fecundity: they contained fewer eggs. One possible explanation for this is that females would be smaller in more dense populations (due to limited resources) and that smaller females would have fewer eggs. This hypothesis was tested and disproved in a 2009 study which showed that female weight had no influence.[3] The actual cause of the density-dependence of fecundity in this organism is still unclear and awaiting further investigation.

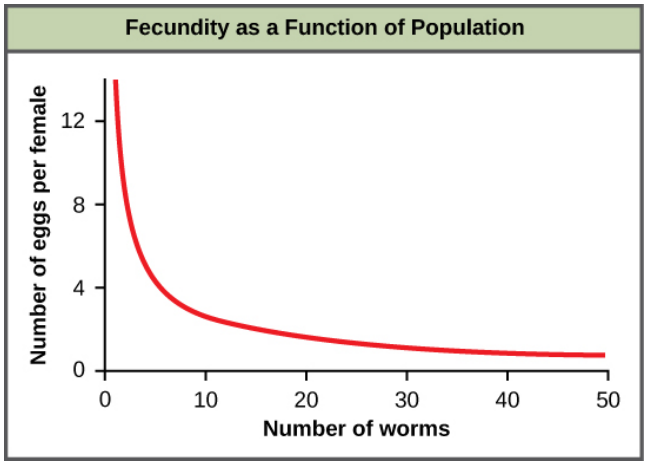


Figure 6.8 In this population of roundworms, fecundity (number of eggs) decreases with population density.[4]

Density-independent Regulation and Interaction with Density-dependent Factors

Many factors, typically physical or chemical in nature (abiotic), influence the mortality of a population regardless of its density, including weather, natural disasters, and pollution. An individual deer may be killed in a forest fire regardless of how many deer happen to be in that area. Its chances of survival are the same whether the population density is high or low. The same holds true for cold winter weather.

In real-life situations, population regulation is very complicated and density-dependent and independent factors can interact. A dense population that is reduced in a density-independent manner by some environmental factor(s) will be able to recover differently than a sparse population. For example, a population of deer affected by a harsh winter will recover faster if there are more deer remaining to reproduce.

---

2. N.A. Croll et al., "The Population Biology and Control of *Ascaris lumbricoides* in a Rural Community in Iran." *Transactions of the Royal Society of Tropical Medicine and Hygiene* 76, no. 2 (1982): 187-197, doi:10.1016/0035-9203(82)90272-3.
3. Martin Walker et al., "Density-Dependent Effects on the Weight of Female *Ascaris lumbricoides* Infections of Humans and its Impact on Patterns of Egg Production." *Parasites & Vectors* 2, no. 11 (February 2009), doi:10.1186/1756-3305-2-11.
4. N.A. Croll et al., "The Population Biology and Control of *Ascaris lumbricoides* in a Rural Community in Iran." *Transactions of the Royal Society of Tropical Medicine and Hygiene* 76, no. 2 (1982): 187-197, doi:10.1016/0035-9203(82)90272-3.

evolution CONNECTION

Why Did the Woolly Mammoth Go Extinct?

(a) (b) (c)

Figure 6.9 The three photos include: (a) 1916 mural of a mammoth herd from the American Museum of Natural History, (b) the only stuffed mammoth in the world, from the Museum of Zoology located in St. Petersburg, Russia, and (c) a one-month-old baby mammoth, named Lyuba, discovered in Siberia in 2007. (credit a: modification of work by Charles R. Knight; credit b: modification of work by "Tanapon"/Flickr; credit c: modification of work by Matt Howry)

It's easy to get lost in the discussion of dinosaurs and theories about why they went extinct 65 million years ago. Was it due to a meteor slamming into Earth near the coast of modern-day Mexico, or was it from some long-term weather cycle that is not yet understood? One hypothesis that will never be proposed is that humans had something to do with it. Mammals were small, insignificant creatures of the forest 65 million years ago, and no humans existed.

Woolly mammoths, however, began to go extinct about 10,000 years ago, when they shared the Earth with humans who were no different anatomically than humans today (**Figure 6.9**). Mammoths survived in isolated island populations as recently as 1700 BC. We know a lot about these animals from carcasses found frozen in the ice of Siberia and other regions of the north. Scientists have sequenced at least 50 percent of its genome and believe mammoths are between 98 and 99 percent identical to modern elephants.

It is commonly thought that climate change and human hunting led to their extinction. A 2008 study estimated that climate change reduced the mammoth's range from 3,000,000 square miles 42,000 years ago to 310,000 square miles 6,000 years ago.[5] It is also well documented that humans hunted these animals. A 2012 study showed that no single factor was exclusively responsible for the extinction of these magnificent creatures.[6] In addition to human hunting, climate change, and reduction of habitat, these scientists demonstrated another important factor in the mammoth's extinction was the migration of humans across the Bering Strait to North America during the last ice age 20,000 years ago.

The maintenance of stable populations was and is very complex, with many interacting factors determining the outcome. It is important to remember that humans are also part of nature. Once we contributed to a species' decline using primitive hunting technology only.

6.4 | Human Population Growth

"Unlike plagues of the dark ages or contemporary diseases we do not understand, the modern plague of overpopulation is soluble by means we have discovered and with resources we possess. What is lacking is not sufficient knowledge of the solution but universal

5. David Nogués-Bravo et al., "Climate Change, Humans, and the Extinction of the Woolly Mammoth." *PLoS Biol* 6 (April 2008): e79, doi:10.1371/journal.pbio.0060079.
6. G.M. MacDonald et al., "Pattern of Extinction of the Woolly Mammoth in Beringia." *Nature Communications* 3, no. 893 (June 2012), doi:10.1038/ncomms1881.

consciousness of the gravity of the problem and education of the billions who are its victim.”

- Martin Luther King, Jr., civil rights leader and Nobel laureate

Concepts of animal population dynamics can be applied to human population growth. Humans are not unique in their ability to alter their environment. For example, beaver dams alter the stream environment where they are built. Humans, however, have the ability to alter their environment to increase its carrying capacity sometimes to the detriment of other species (e.g., via artificial selection for crops that have a higher yield). Earth’s human population is growing rapidly, to the extent that some worry about the ability of the earth’s environment to sustain this population, as long-term exponential growth carries the potential risks of famine, disease, and large-scale death.

Although humans have increased the carrying capacity of their environment, the technologies used to achieve this transformation have caused unprecedented changes to Earth’s environment, altering ecosystems to the point where some may be in danger of collapse. The depletion of the ozone layer, erosion due to acid rain, and damage from global climate change are caused by human activities. The ultimate effect of these changes on our carrying capacity is unknown. As some point out, it is likely that the negative effects of increasing carrying capacity will outweigh the positive ones—the carrying capacity of the world for human beings might actually decrease.

The world’s human population is currently experiencing exponential growth even though human reproduction is far below its biotic potential (**Figure 6.10**). To reach its biotic potential, all females would have to become pregnant every nine months or so during their reproductive years. Also, resources would have to be such that the environment would support such growth. Neither of these two conditions exists. In spite of this fact, human population is still growing exponentially.

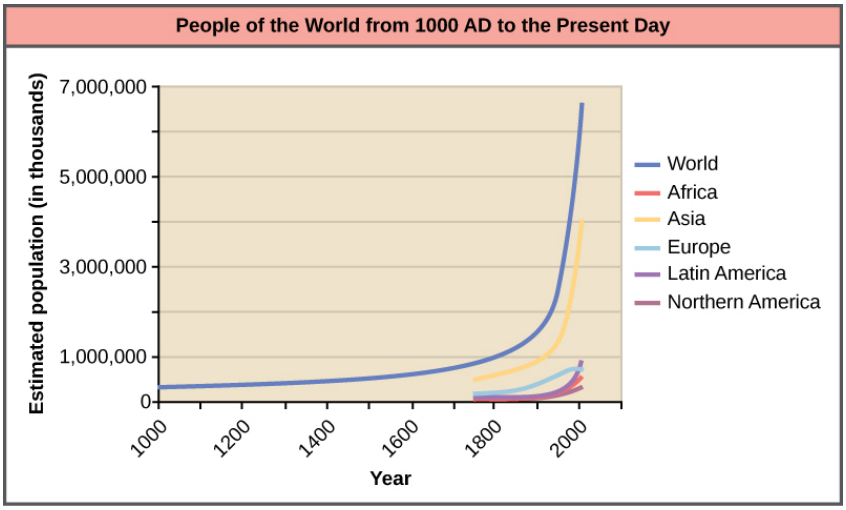


Figure 6.10 Human population growth since 1000 AD is exponential (dark blue line). Notice that while the population in Asia (yellow line), which has many economically underdeveloped countries, is increasing exponentially, the population in Europe (light blue line), where most of the countries are economically developed, is growing much more slowly.

A consequence of exponential human population growth is the time that it takes to add a particular number of humans to the Earth is becoming shorter. **Figure 6.11** shows that 130 years were necessary to add 1 billion humans in 1930, but it only took 24 years to add two billion people between 1975 and 1999. As already discussed, at some point it would appear that our ability to increase our carrying capacity indefinitely on a finite world is uncertain. Without new technological advances, the human growth rate has been predicted to slow in the coming decades. However, the population will still be increasing and the threat of overpopulation remains.

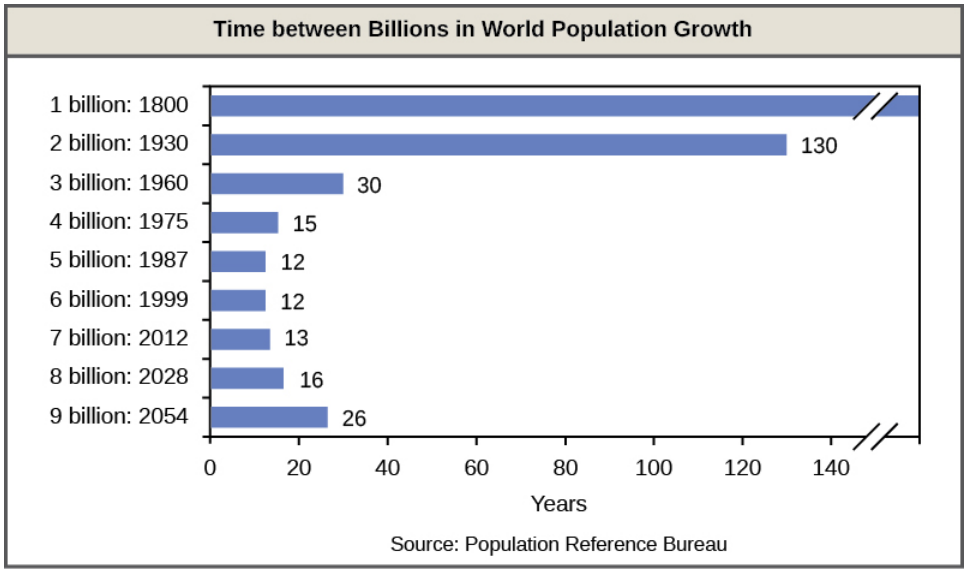


Figure 6.11 The time between the addition of each billion human beings to Earth decreases over time. (credit: modification of work by Ryan T. Cragun)

Overcoming Density-Dependent Regulation

Humans are unique in their ability to alter their environment with the conscious purpose of increasing its carrying capacity. This ability is a major factor responsible for human population growth and a way of overcoming density-dependent growth regulation. Much of this ability is related to human intelligence, society, and communication. Humans can construct shelter to protect them from the elements and have developed agriculture and domesticated animals to increase their food supplies. In addition, humans use language to communicate this technology to new generations, allowing them to improve upon previous accomplishments.

Other factors in human population growth are migration and public health. Humans originated in Africa, but have since migrated to nearly all inhabitable land on the Earth. Public health, sanitation, and the use of antibiotics and vaccines have decreased the ability of infectious disease to limit human population growth. In the past, diseases such as the bubonic plaque of the fourteenth century killed between 30 and 60 percent of Europe's population and reduced the overall world population by as many as 100 million people. Today, the threat of infectious disease, while not gone, is certainly less severe. According to the World Health Organization, global death from infectious disease declined from 16.4 million in 1993 to 14.7 million in 1992. To compare to some of the epidemics of the past, the percentage of the world's population killed between 1993 and 2002 decreased from 0.30 percent of the world's population to 0.24 percent. Thus, it appears that the influence of infectious disease on human population growth is becoming less significant.

Age Structure, Population Growth, and Economic Development

The age structure of a population is an important factor in population dynamics. **Age structure** is the proportion of a population at different age ranges. Age structure allows better prediction of population growth, plus the ability to associate this growth with the level of economic development in the region. Countries with rapid growth have a pyramidal shape in their age structure diagrams, showing a preponderance of younger individuals, many of whom are of reproductive age or will be soon (**Figure 6.12**). This pattern is most often observed in underdeveloped countries where individuals do not live to old age because of less-than-optimal living conditions. Age structures of areas with slow growth, including developed countries such as the United States, have a parabola shape structure (stage 3) (**Figure 6.12**), with many fewer young and reproductive-aged individuals and a greater proportion of older individuals. Other developed countries, such as Italy, have zero population growth. The age structure of these populations is more conical, with an even greater percentage of middle-aged and older individuals. The actual growth rates in different countries are shown in **Figure 6.13**, with the highest rates tending to be in the less economically developed countries of Africa and Asia. Lastly, both fertility rates and age structure interact to influence population growth rates. High fertility rates can increase population growth because more offspring are produced per individual, whereas a pyramid-like age structure increases growth rates because more individuals in the population are producing offspring.

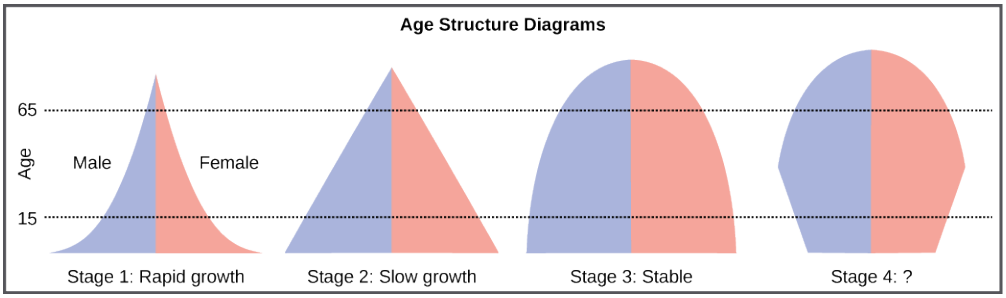


Figure 6.12 Typical age structure diagrams are shown. The rapid growth diagram narrows to a point, indicating that the number of individuals decreases rapidly with age. In the slow growth model, the number of individuals decreases steadily with age. Stable population diagrams are rounded on the top, showing that the number of individuals per age group decreases gradually, and then increases for the older part of the population.

Age structure diagrams for rapidly growing, slow growing and stable populations are shown in stages 1 through 3. What type of population change do you think stage 4 represents? If you guessed decreasing, you are correct.

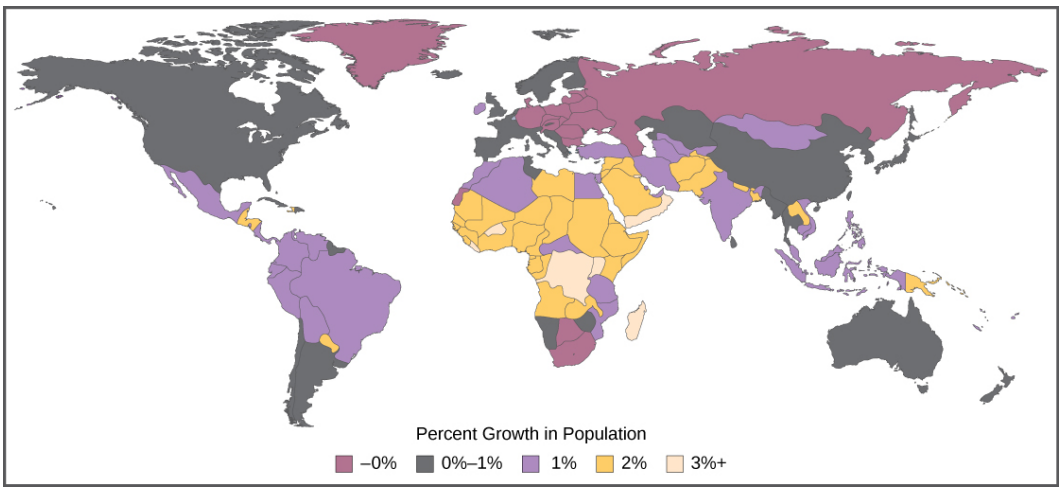


Figure 6.13 The percent growth rate of population in different countries is shown. Notice that the highest growth is occurring in less economically developed countries in Africa and Asia.

Long-Term Consequences of Exponential Human Population Growth

Many dire predictions have been made about the world's population leading to a major crisis called the "population explosion." In the 1968 book *The Population Bomb*, biologist Dr. Paul R. Ehrlich wrote, "The battle to feed all of humanity is over. In the 1970s hundreds of millions of people will starve to death in spite of any crash programs embarked upon now. At this late date nothing can prevent a substantial increase in the world death rate."[7] While many critics view this statement as an exaggeration, the laws of exponential population growth are still in effect, and unchecked human population growth cannot continue indefinitely.

Efforts to control population growth led to the **one-child policy** in China, which used to include more severe consequences, but now imposes fines on urban couples who have more than one child. Due to the fact that some couples wish to have a male heir, many Chinese couples continue to have more than one child. The policy itself, its social impacts, and the effectiveness of limiting overall population growth are controversial. In spite of population control policies, the human population continues to grow. At some point the food supply may run out because of the subsequent need to produce more and more food to feed our population. The United Nations estimates that future world population growth may vary from 6 billion (a decrease) to 16 billion people by the year 2100. There is no way to know whether human population growth will moderate to the point where the crisis described by Dr. Ehrlich will be averted.

Another result of population growth is the endangerment of the natural environment. Many countries have attempted to reduce the human impact on climate change by reducing their emission of the greenhouse gas carbon dioxide. However, these treaties have not been ratified by every country, and many underdeveloped countries trying to improve their economic

7. Paul R. Erlich, prologue to *The Population Bomb*, (1968; repr., New York: Ballantine, 1970).

condition may be less likely to agree with such provisions if it means slower economic development. Furthermore, the role of human activity in causing climate change has become a hotly debated socio-political issue in some developed countries, including the United States. Thus, we enter the future with considerable uncertainty about our ability to curb human population growth and protect our environment.

7 | COMMUNITY ECOLOGY

> “The outstanding scientific discovery of the twentieth century is not television, or radio, but rather the complexity of the land organism. Only those who know the most about it can appreciate how little we know about it. The last word in ignorance is the man who says of an animal or plant: "What good is it?" If the land mechanism as a whole is good, then every part is good, whether we understand it or not.”

- Aldo Leopold, A Sand County Almanac

Populations rarely, if ever, live in isolation from populations of other species. In most cases, numerous species share a habitat. The interactions between these populations play a major role in regulating population growth and abundance. All populations occupying the same habitat form a **community**: all populations inhabiting a specific area at the same time. The number of species occupying the same habitat and their relative abundance is known as species diversity. Areas with low diversity, such as the glaciers of Antarctica, still contain a wide variety of living things, whereas the diversity of tropical rainforests is so great that it cannot be counted. Ecology is studied at the community level to understand how species interact with each other and what processes determine the patterns of species coexistence, diversity, and distributions that we see in nature Any interactions between two or more species is referred to globally as **interspecific interactions** , and there are specific terms for some of unique interactions that will be discussed throughout this section. These unique interactions can have a positive effect (+), a negative effect (-) or a neutral effect (0) for the individual of the species.

Niche

In a community, no two species can have exactly the same ecological requirement and persist together in equilibrium. In order for a species to survive in an area, to obtain nutrients and energy and to avoid predators, a certain combination of ecological variables are necessary. These ecological variables include the resources available (light, nutrients, food and habitat), environmental characteristics (temperature and water availability) and the interactions with other organisms. These ecological requirements and interactions are used to define an organisms **niche**. For example, the Black-tailed Prairie Dog (*Cynomys ludovicianus*) **Figure 7.1** is mostly found in short grass prairie. They can live within a broad range of temperature and moisture, specific soil types, and forage on specific species of grasses. These characteristics, other variables and interspecific interactions define their niche.

Figure 7.1 Black-Tailed Prairie Dog The Black-tailed Prairie Dog has a unique niche. They eat grass and are eaten by coyotes, black-footed ferrets and other carnivores. The prairie dog lives in burrows and this activity produces bare soil areas where seeds can germinate without being shaded by trees. The burrows once abandoned provide nesting sites for burrowing owls.(photo by D. A. Rintoul)

Predation and Herbivory

Perhaps the classical example of species interaction is predation: the hunting of prey by its predator where the predator is positively affected (+) and the prey is negatively affected (-). Nature shows on television highlight the drama of one living organism killing another. Populations of predators and prey in a community are not constant over time: in most cases, they vary in cycles that appear to be related. The most often cited example of predator-prey dynamics is seen in the cycling of the lynx (predator) and the snowshoe hare (prey), using nearly 200 year-old trapping data from North American forests (**Figure 7.2**). This cycle of predator and prey repeats itself approximately every 10 years, with the predator population lagging 1–2 years behind that of the prey population. As the hare numbers increase, there is more food available for the lynx, allowing the lynx population to increase as well. When the lynx population grows to a threshold level, however, they kill so many hares that hare population begins to decline, followed by a decline in the lynx population because of scarcity of food. When the lynx population is low, the hare population size begins to increase due, at least in part, to low predation pressure, starting the cycle anew.

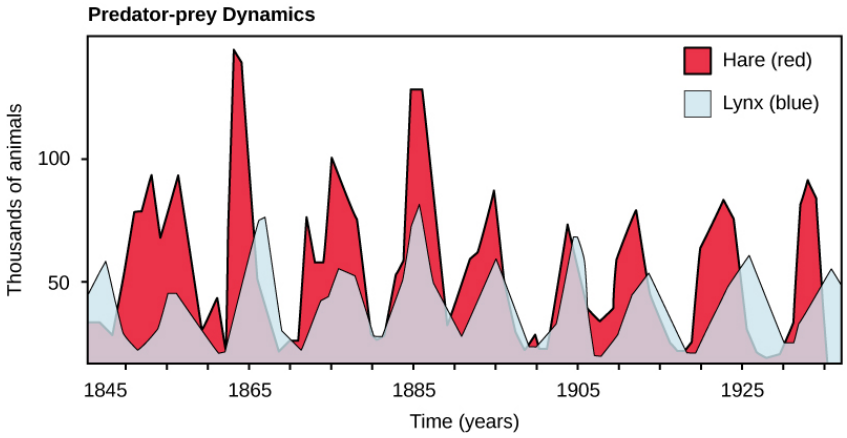


Figure 7.2 The cycling of lynx and snowshoe hare populations in Northern Ontario is an example of predator-prey dynamics.

The idea that the population cycling of the two species is entirely controlled by predation models has come under question. More recent studies have pointed to undefined density-dependent factors as being important in the cycling, in addition to

predation. One possibility is that the cycling is inherent in the hare population due to density-dependent effects such as lower fecundity (maternal stress) caused by crowding when the hare population gets too dense. The hare cycling would then induce the cycling of the lynx because it is the lynxes' major food source. In addition, many populations of annual plants show periodic cycles that are completely unrelated to any predator or herbivore population. The more we study communities, the more complexities we find, allowing ecologists to derive more accurate and sophisticated models of population dynamics.

Herbivory describes the consumption of plants by insects and other animals, and it is another interspecific relationship that affects populations. Unlike animals, most plants cannot outrun predators or use mimicry to hide from hungry animals. Some plants have developed mechanisms to defend against herbivory. Other species have developed mutualistic relationships; for example, herbivory provides a mechanism of seed distribution that aids in plant reproduction.

Defense Mechanisms against Predation and Herbivory

The study of communities must consider evolutionary forces that act on the members of the various populations contained within it. Species are not static, but slowly changing and adapting to their environment by natural selection and other evolutionary forces. Species have evolved numerous mechanisms to escape predation and herbivory. These defenses may be mechanical, chemical, physical, or behavioral.

Mechanical defenses, such as the presence of thorns on plants or the hard shell on turtles, discourage animal predation and herbivory by causing physical pain to the predator or by physically preventing the predator from being able to eat the prey. Chemical defenses are produced by many animals as well as plants, such as the foxglove which is extremely toxic when eaten. **Figure 7.3** shows some organisms' defenses against predation and herbivory.

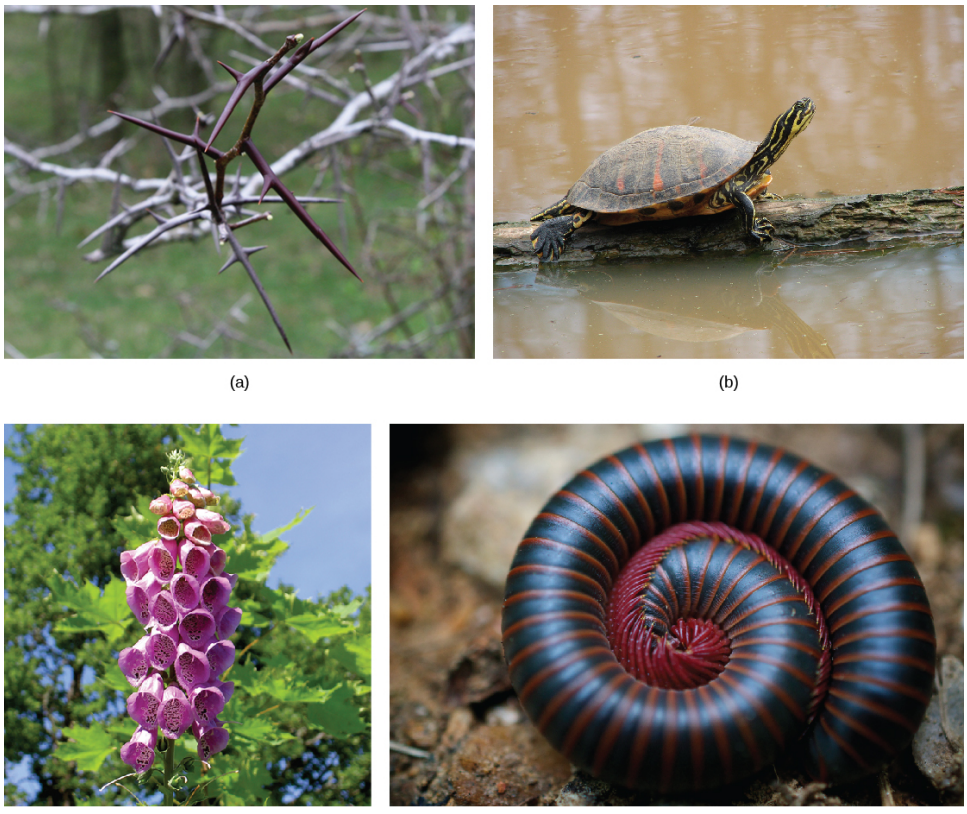


Figure 7.3 The (a) honey locust tree (*Gleditsia triacanthos*) uses thorns, a mechanical defense, against herbivores, while the (b) Florida red-bellied turtle (*Pseudemys nelsoni)* uses its shell as a mechanical defense against predators. (c) Foxglove (*Digitalis* sp.) uses a chemical defense: toxins produced by the plant can cause nausea, vomiting, hallucinations, convulsions, or death when consumed. (d) The North American millipede (*Narceus americanus*) uses both mechanical and chemical defenses: when threatened, the millipede curls into a defensive ball and produces a noxious substance that irritates eyes and skin. (credit a: modification of work by Huw Williams; credit b: modification of work by "JamieS93"/Flickr; credit c: modification of work by Philip Jägenstedt; credit d: modification of work by Cory Zanker)

Many species use their body shape and coloration to avoid being detected by predators. The crab spider has the coloration

and body shape of a flower petal which makes it very hard to see when stationary against a background of real a real flower (**Figure 7.4a**). In another example, the chameleon can change its color to match its surroundings (**Figure 7.4b**). Both of these are examples of **camouflage**, or avoiding detection by blending in with the background.

(a) (b)

Figure 7.4 (a) The crab spider (a) and the chameleon (b) use body shape and/or coloration to deceive potential prey, and to prevent detection by predators. (credit a: photograph by David A. Rintoul; credit b: modification of work by Frank Vassen)

Some species use coloration as a way of warning predators that they are not good to eat. For example, the cinnabar moth caterpillar, the fire-bellied toad, and many species of beetle have bright colors that warn of a foul taste, the presence of toxic chemical, and/or the ability to sting or bite, respectively. Predators that ignore this coloration and eat the organisms will experience their unpleasant taste or presence of toxic chemicals and learn not to eat them in the future. This type of defensive mechanism is called **aposematic coloration**, or **warning coloration** (**Figure 7.5**).

(a) (b)

Figure 7.5 (a) The strawberry poison dart frog (*Oophaga pumilio*) uses aposematic coloration to warn predators that it is toxic, while the (b) striped skunk (*Mephitis mephitis*) uses aposematic coloration to warn predators of the unpleasant odor it produces. (credit a: modification of work by Jay Iwasaki; credit b: modification of work by Dan Dzurisin)

While some predators learn to avoid eating certain potential prey because of their coloration, other species have evolved mechanisms to mimic this coloration to avoid being eaten, even though they themselves may not be unpleasant to eat or contain toxic chemicals. In **Batesian mimicry**, a harmless species imitates the warning coloration of a harmful one. Assuming they share the same predators, this coloration then protects the harmless ones, even though they do not have the same level of physical or chemical defenses against predation as the organism they mimic. Many insect species mimic the coloration of wasps or bees, which are stinging, venomous insects, thereby discouraging predation (**Figure 7.6**).

(a) (b)

Figure 7.6 Batesian mimicry occurs when a harmless species mimics the coloration of a harmful species, as is seen with the (a) bumblebee and (b) bee-like robber fly. (credit a, b: modification of work by Cory Zanker)

In Müllerian mimicry, multiple species share the same warning coloration, but all of them actually have defenses. **Figure 7.7** shows a variety of foul-tasting butterflies with similar coloration.

Figure 7.7 Several unpleasant-tasting *Heliconius* butterfly species share a similar color pattern, an example of Müllerian mimicry. (credit: Joron M, Papa R, Beltrán M, Chamberlain N, Mavárez J, et al.)

Competition

Resources are often limited within a habitat and multiple species may compete to obtain them. All species have an ecological niche in the ecosystem, which describes how they acquire the resources they need and how they interact with other species in the community. So, the plants in a garden are competing with each other for soil nutrients, and water. This competition between the different species is called interspecific competition. The overall effect on both species is negative because either one of the species would do better if the other species is not present. So, how do species reduce the overall negative effects of direct competition?

Competitive Exclusion Principle

The **competitive exclusion principle** states that two species cannot occupy the same niche in a habitat and stably coexist. In other words, different species cannot coexist in a community if they are competing for all the same resources. An example of

this principle is shown in **Figure 7.8**, with two protozoan species, *Paramecium aurelia* and *Paramecium caudatum*. When grown individually in the laboratory, they both thrive. But when they are placed together in the same test tube (habitat), *P. aurelia* outcompetes *P. caudatum* for food, leading to the latter's eventual extinction.

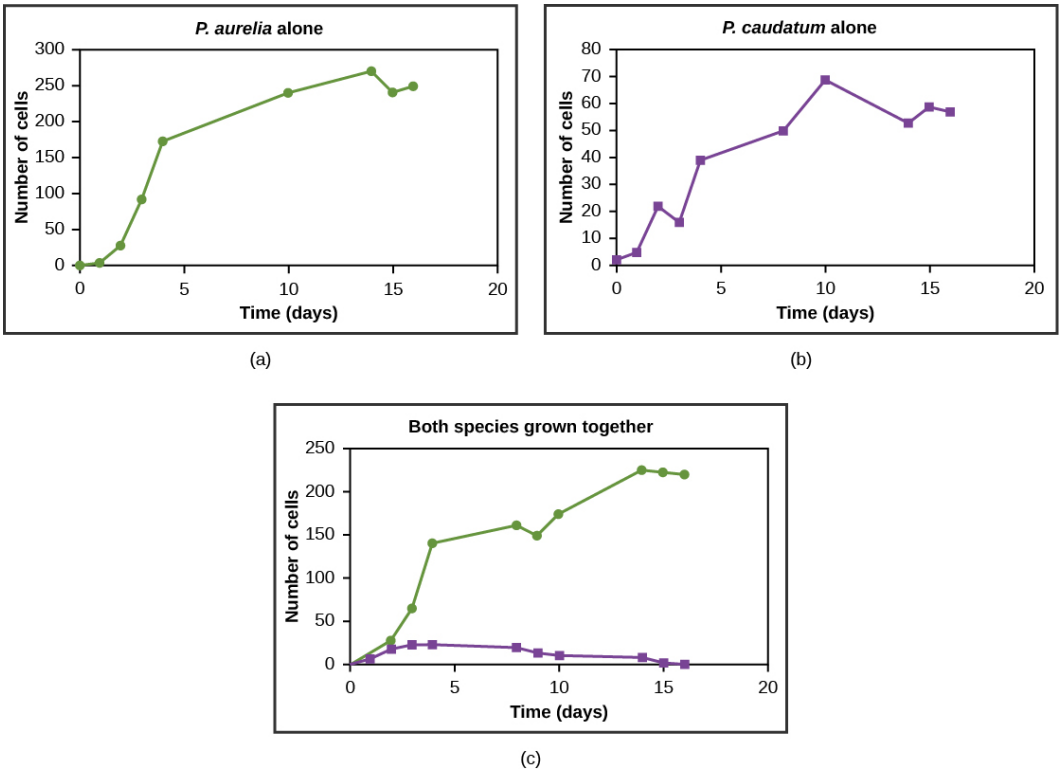


Figure 7.8 *Paramecium aurelia* and *Paramecium caudatum* grow well individually, but when they compete for the same resources, the *P. aurelia* outcompetes the *P. caudatum*.

Resource patitioning

This exclusion may be avoided if a population evolves to make use of a different resource, a different area of the habitat, or feeds during a different time of day, called resource partitioning. The two organisms are then said to occupy different niches. These organisms coexist by minimizing direct competition. The anole lizards found on a single island are a good example of resource partitioning **Figure 7.9** because it shows the effects of how natural selection has driven the evolution of different species in order to reduce competition.

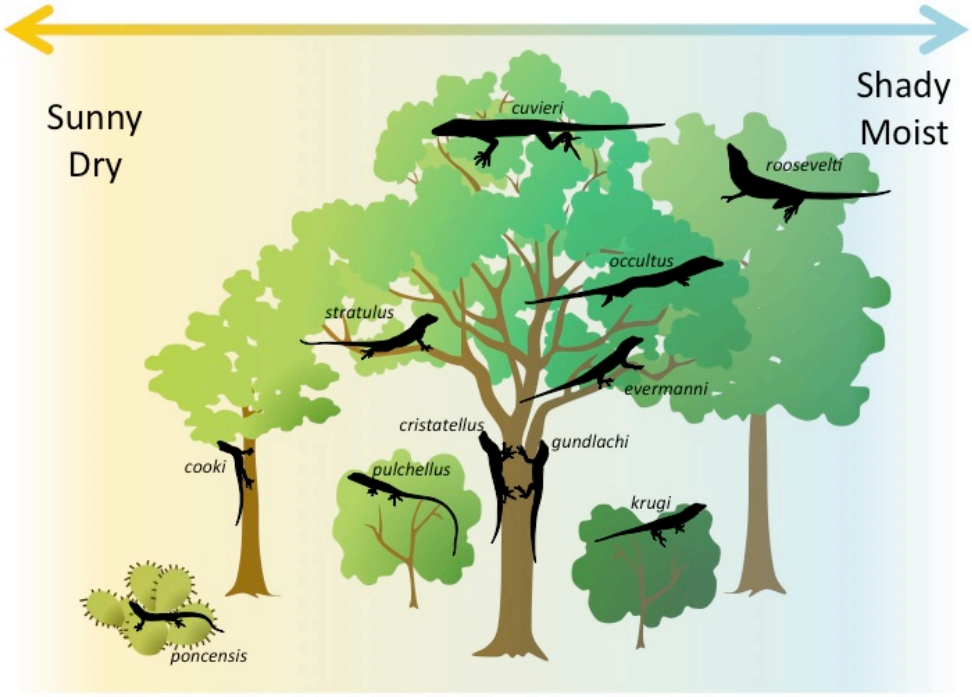


Figure 7.9 This figure shows resource partitioning among 11 species of anole lizards found on the island of Puerto Rico. Each species occupies a different type or elevation of vegetation. The habitat is further partitioned by the amount of sunlight and moisture available. Image by Eva Horne modified from (Williams, E.E. 1983. Ecomorphs, faunas, island size, and diverse end points in island radiations of Anolis. In Lizard Ecology: Studies of a Model Organism. Eds. R.B. Huey, E.R. Pianka, and T.W. Schoener. Harvard University Press).

Symbiosis

Symbiotic relationships, or **symbioses** (plural), are close interactions between individuals of different species over an extended period of time that impact the abundance and distribution of the associating populations. Symbiosis is a greek word meaning “living together”. At least one of the species is dependent upon the other (they are not free-living), and often both species are dependent. Most scientists accept this definition, but some restrict the term to only those species that are mutualistic, where both individuals benefit from the interaction. In this discussion, the first broader definition will be used.

Commensalism

A **commensal** relationship occurs when one species benefits (+) from the close, prolonged interaction, while the other neither benefits nor is harmed (0). Birds nesting in trees provide an example of a commensal relationship (**Figure 7.10**). The tree is not harmed by the presence of the nest among its branches. The nests are light and produce little strain on the structural integrity of the branch, and most of the leaves, which the tree uses to get energy by photosynthesis, are above the nest so they are unaffected. The bird, on the other hand, benefits greatly. If the bird had to nest in the open, its eggs and young would be vulnerable to predators. Another example of a commensal relationship is the clown fish and the sea anemone. The sea anemone is not harmed by the fish, and the fish gains protection from predators who would be stung upon nearing the sea anemone.

Figure 7.10 The southern masked-weaver bird is starting to make a nest in a tree in Zambezi Valley, Zambia. This is an example of a commensal relationship, in which one species (the bird) benefits, while the other (the tree) neither benefits nor is harmed. (credit: "Hanay"/Wikimedia Commons)

Mutualism

A second type of species interaction is called **mutualism**, where two species benefit from their interaction (+,+). For example, termites have a mutualistic relationship with protozoa that live in the insect's gut (**Figure 7.11a**). The termite benefits from the ability of bacterial symbionts within the protozoa to digest cellulose. The termite itself cannot do this, and without the protozoa, it would not be able to obtain energy from its food (cellulose from the wood it chews and eats). The protozoa and the bacterial symbionts benefit by having a protective environment and a constant supply of food from the wood chewing actions of the termite. Lichens have a mutualistic relationship between fungus and photosynthetic algae or bacteria (**Figure 7.11b**). As these symbionts grow together, the glucose produced by the algae provides nourishment for both organisms, whereas the physical structure of the lichen protects the algae from the elements and makes certain nutrients in the atmosphere more available to the algae. Another example of mutualism is the interaction between a plant and its insect pollinator. The plant benefits from having its pollen transferred to another individual to carry out reproduction and the insect benefits from nectar or other reward provided by the plant. This is an example of a non-symbiotic mutualism as both the plant and the insect are free-living organisms.

(a) (b)

Figure 7.11 (a) Termites form a mutualistic relationship with symbiotic protozoa in their guts, which allow both organisms to obtain energy from the cellulose the termite consumes. (b) Lichen is a fungus that has symbiotic photosynthetic algae living inside its cells. (credit a: modification of work by Scott Bauer, USDA; credit b: modification of work by Cory Zanker)

Parasitism

A **parasite** is an organism that lives in or on another living organism and derives nutrients from it. In this relationship, the parasite benefits (+), but the organism being fed upon, the **host** is harmed (-). The host is usually weakened by the parasite as it siphons resources the host would normally use to maintain itself. The parasite, however, is unlikely to kill the host, especially not quickly, because this would allow no time for the organism to complete its reproductive cycle by spreading to another host.

The reproductive cycles of parasites are often very complex, sometimes requiring more than one host species. A tapeworm is a parasite that causes disease in humans when contaminated, undercooked meat such as pork, fish, or beef is consumed (**Figure 7.12**). The tapeworm can live inside the intestine of the host for several years, benefiting from the food the host is bringing into its gut by eating, and may grow to be over 50 ft long by adding segments. The parasite moves from species to species in a cycle, making two hosts necessary to complete its life cycle. Another common parasite is *Plasmodium falciparum*, the protozoan cause of malaria, a significant disease in many parts of the world. Living in human liver and red blood cells, the organism reproduces asexually in the gut of blood-feeding mosquitoes to complete its life cycle. Thus malaria is spread from human to human by mosquitoes, one of many arthropod-borne infectious diseases.

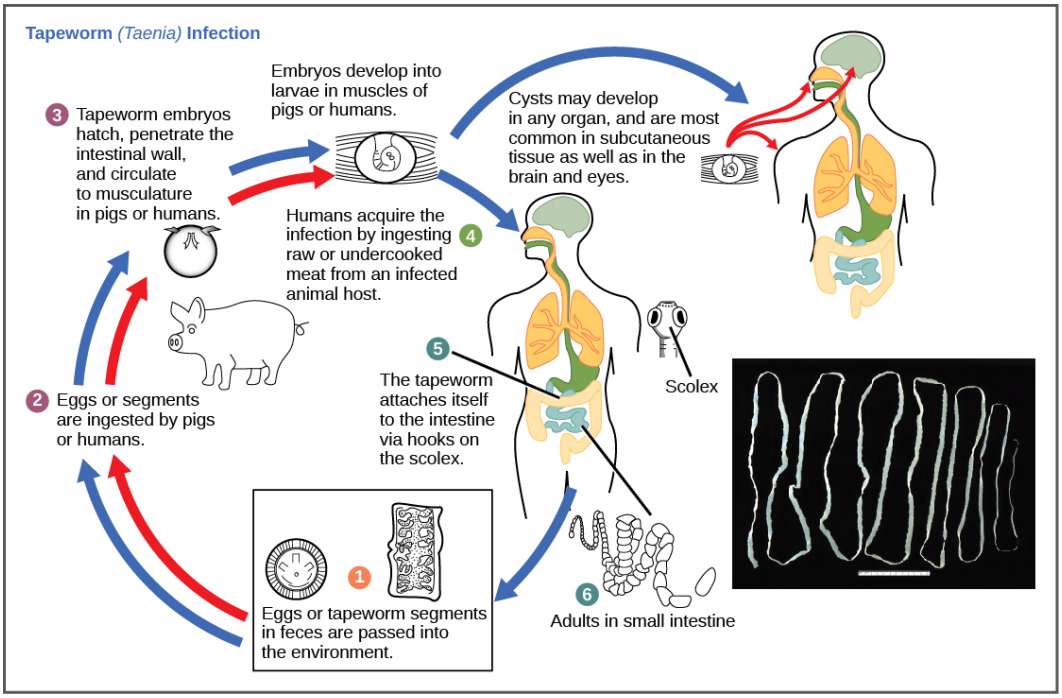


Figure 7.12 This diagram shows the life cycle of a pork tapeworm (*Taenia solium*), a human worm parasite. (credit: modification of work by CDC)

Amensalism

Another type of interaction classified by biologists and ecologists is amensalism. Amensalism is any interaction between individuals of different species in which one individual is harmed (-) while the other individual is not affected (0). For example, as you walk down a side walk on a rainy day you step on an earthworm; the earthworm is negatively affected, and you are not affected. Amensalism occurs among micro-organisms when microbe species A releases chemicals that have a negative effect on microbe species B, but B has neither a positive nor negative effect on A.

Coevolution

When the genetic change in one species causes a subsequent change in the genetic structure of another species, this is called coevolution. In a community, all the interacting species have the potential to influence one another, and in a sense they are all evolving together. However, coevolution can only describe genetic changes in interacting species if scientists can demonstrate that specific interactions result in reciprocal adaptations. For example, a species of plant may rely solely on one species of insect for pollination and that one species of insect may only consume nectar from that one flower. Many of the above examples of species interactions do not fit the strict definition of coevolution, but one can not argue that these species are evolving in response to one another and their environment.

Characteristics of Communities

Communities are complex entities that can be characterized by their structure (the types and numbers of species present) and dynamics (how communities change over time). Understanding community structure and dynamics enables community ecologists to manage ecosystems more effectively.

Foundation Species

Foundation species are considered the “base” or “bedrock” of a community, having the greatest influence on its overall

structure. They are usually the primary producers: organisms that bring most of the energy into the community. Kelp, brown algae, is a foundation species, forming the basis of the kelp forests off the coast of California.

Foundation species may physically modify the environment to produce and maintain habitats that benefit the other organisms that use them. An example is the photosynthetic corals of the coral reef (**Figure 7.13**). Corals themselves are not photosynthetic, but harbor symbionts within their body tissues (dinoflagellates called zooxanthellae) that perform photosynthesis; this is another example of a mutualism. The exoskeletons of living and dead coral make up most of the reef structure, which protects many other species from waves and ocean currents.

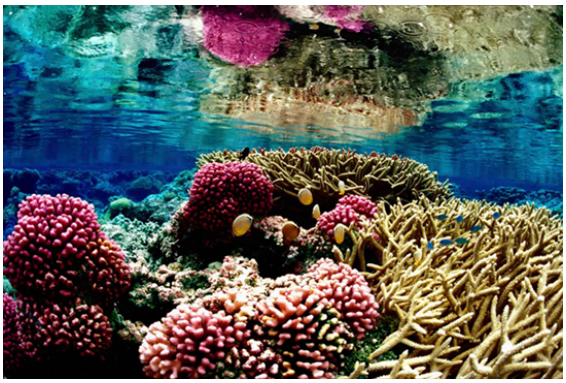


Figure 7.13 Coral is the foundation species of coral reef ecosystems. (credit: Jim E. Maragos, USFWS)

Biodiversity, Species Richness, and Relative Species Abundance

Biodiversity describes a community's biological complexity: it is measured by the number of different species (species richness) in a particular area and their relative abundance (species evenness). The area in question could be a habitat, a biome, or the entire biosphere. Species richness is the term that is used to describe the number of species living in a habitat or biome. Species richness varies across the globe (**Figure 7.14**). One factor in determining species richness is latitude, with the greatest species richness occurring in ecosystems near the equator, which often have warmer temperatures, large amounts of rainfall, and low seasonality. The lowest species richness occurs near the poles, which are much colder, drier, and thus less conducive to life in Geologic time (time since glaciations). The predictability of climate or productivity is also an important factor. Other factors influence species richness as well. Relative species abundance is the number of individuals in a species relative to the total number of individuals in all species within a habitat, ecosystem, or biome. Foundation species often have the highest relative abundance of species.

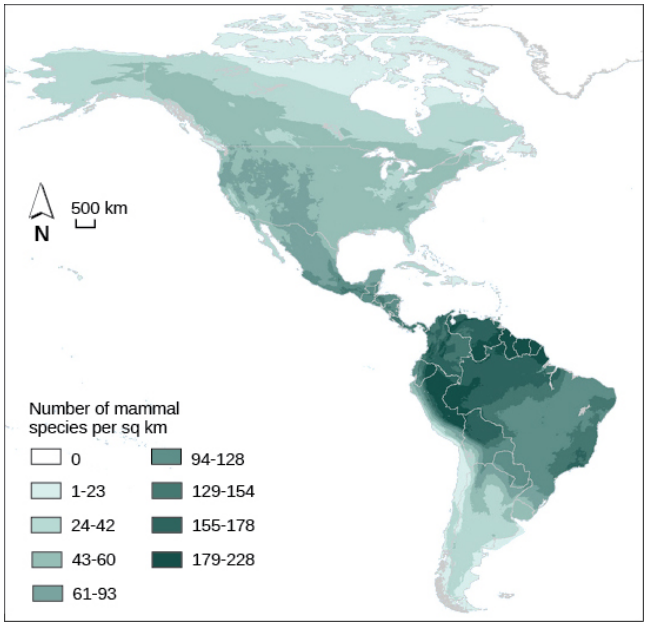


Figure 7.14 The greatest species richness for mammals in North and South America is associated with the equatorial latitudes. (credit: modification of work by NASA, CIESIN, Columbia University)

Keystone Species

A keystone species is one species that has a disproportionately large effect on community structure relative to its biomass or abundance. The intertidal sea star, *Pisaster ochraceus*, of the northwestern United States is a keystone species (**Figure 7.15**). Studies have shown that when this organism is removed from communities, populations of their natural prey (mussels) increase, completely altering the species composition and reducing biodiversity. Another keystone species is the banded tetra, a fish in tropical streams, which supplies nearly all of the phosphorus, a necessary inorganic nutrient, to the rest of the community. If these fish were to become extinct, the community would be greatly affected.

Figure 7.15 The *Pisaster ochraceus* sea star is a keystone species. (credit: Jerry Kirkhart)

everyday CONNECTION

Invasive Species

Invasive species are non-native organisms that, when introduced to an area out of their native range, threaten the ecosystem balance of that habitat. Many such species exist in the United States, as shown in **Figure 7.16**. Whether enjoying a forest hike, taking a summer boat trip, or simply walking down an urban street, you have likely encountered an invasive species.

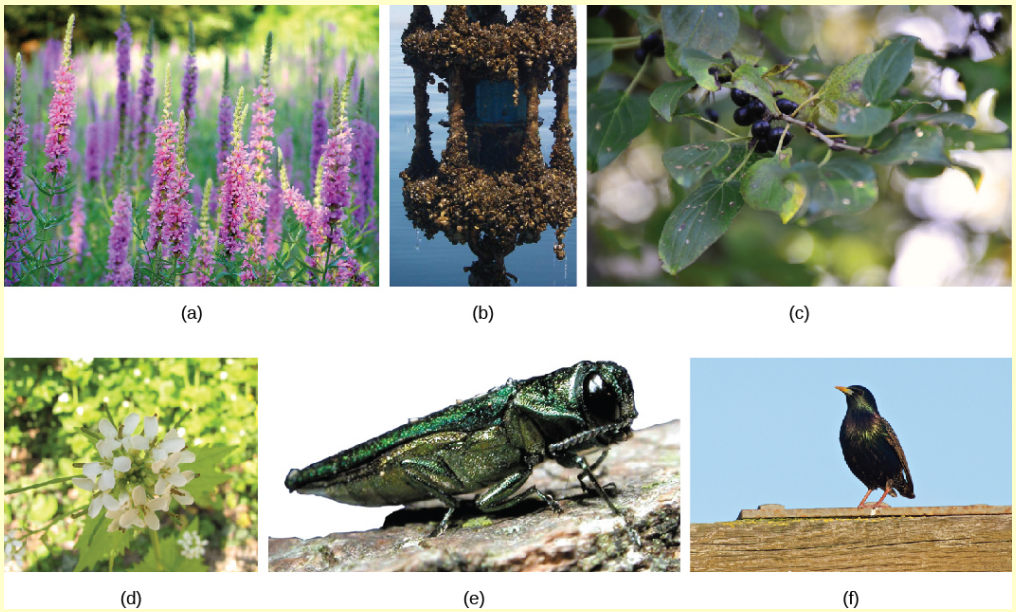


Figure 7.16 In the United States, invasive species like (a) purple loosestrife (*Lythrum salicaria*) and the (b) zebra mussel (*Dreissena polymorpha*) threaten certain aquatic ecosystems. Some forests are threatened by the spread of (c) common buckthorn (*Rhamnus cathartica*), (d) garlic mustard (*Alliaria petiolata*), and (e) the emerald ash borer (*Agrilus planipennis*). The (f) European starling (*Sturnus vulgaris*) may compete with native bird species for nest holes. (credit a: modification of work by Liz West; credit b: modification of work by M. McCormick, NOAA; credit c: modification of work by E. Dronkert; credit d: modification of work by Dan Davison; credit e: modification of work by USDA; credit f: modification of work by Don DeBold)

One of the many recent proliferations of an invasive species concerns the growth of Asian carp populations. Asian carp were introduced to the United States in the 1970s by fisheries and sewage treatment facilities that used the fish's excellent filter feeding capabilities to clean their ponds of excess plankton. Some of the fish escaped, however, and by the 1980s they had colonized many waterways of the Mississippi River basin, including the Illinois and Missouri Rivers.

Voracious eaters and rapid reproducers, Asian carp may outcompete native species for food, potentially leading to their extinction. For example, black carp are voracious eaters of native mussels and snails, limiting this food source for native fish species. Silver carp eat plankton that native mussels and snails feed on, reducing this food source by a different alteration of the food web. In some areas of the Mississippi River, Asian carp species have become the most predominant, effectively outcompeting native fishes for habitat. In some parts of the Illinois River, Asian carp constitute 95 percent of the community's biomass. Although edible, the fish is bony and not a desired food in the United States. Moreover, their presence threatens the native fish and fisheries of the Great Lakes, which are important to local economies and recreational anglers. Asian carp have even injured humans. The fish, frightened by the sound of approaching motorboats, thrust themselves into the air, often landing in the boat or directly hitting the boaters.

The Great Lakes and their prized salmon and lake trout fisheries are also being threatened by these invasive fish. Asian carp have already colonized rivers and canals that lead into Lake Michigan. One infested waterway of particular importance is the Chicago Sanitary and Ship Channel, the major supply waterway linking the Great Lakes to the Mississippi River. To prevent the Asian carp from leaving the canal, a series of

electric barriers have been successfully used to discourage their migration; however, the threat is significant enough that several states and Canada have sued to have the Chicago channel permanently cut off from Lake Michigan. Local and national politicians have weighed in on how to solve the problem, but no one knows whether the Asian carp will ultimately be considered a nuisance, like other invasive species such as the water hyacinth and zebra mussel, or whether it will be the destroyer of the largest freshwater fishery of the world.

The issues associated with Asian carp show how population and community ecology, fisheries management, and politics intersect on issues of vital importance to the human food supply and economy. Socio-political issues like this make extensive use of the sciences of population ecology (the study of members of a particular species occupying a particular area known as a habitat) and community ecology (the study of the interaction of all species within a habitat).

Community Dynamics

Community dynamics are the changes in community structure and composition over time. Sometimes these changes are induced by environmental disturbances such as volcanoes, earthquakes, storms, fires, and climate change. Communities with a stable structure are said to be at equilibrium. Following a disturbance, the community may or may not return to the equilibrium state.

Succession describes the sequential appearance and disappearance of species in a community over time. In **primary succession**, newly exposed or newly formed land is colonized by living things; in **secondary succession**, part of an ecosystem is disturbed and remnants of the previous community remain.

Primary Succession and Pioneer Species

Primary succession occurs when new land is formed or rock is exposed: for example, following the eruption of volcanoes, such as those on the Big Island of Hawaii. As lava flows into the ocean, new land is continually being formed. On the Big Island, approximately 32 acres of land is added each year. First, weathering and other natural forces break down the substrate enough for the establishment of certain hearty plants and lichens with few soil requirements, known as **pioneer species** (Figure 7.17). These species help to further break down the mineral rich lava into soil where other, less hardy species will grow and eventually replace the pioneer species. In addition, as these early species grow and die, they add to an ever-growing layer of decomposing organic material and contribute to soil formation. Over time the area will reach an equilibrium state, with a set of organisms quite different from the pioneer species.

Figure 7.17 During primary succession in lava on Maui, Hawaii, succulent plants are the pioneer species. (credit: Forest and Kim Starr)

Secondary succession

A classic example of secondary succession occurs in oak and hickory forests cleared by wildfire (Figure 7.18). Wildfires will burn most vegetation and kill those animals unable to flee the area. Their nutrients, however, are returned to the ground in the form of ash. Thus, even when areas are devoid of life due to severe fires, the area will soon be ready for new life to take hold.

Before the fire, the vegetation was dominated by tall trees with access to the major plant energy resource: sunlight. Their height gave them access to sunlight while also shading the ground and other low-lying species. After the fire, though, these trees are no longer dominant. Thus, the first plants to grow back are usually annual plants followed within a few years

by quickly growing and spreading grasses and other pioneer species. Due to, at least in part, changes in the environment brought on by the growth of the grasses and other species, over many years, shrubs will emerge along with small pine, oak, and hickory trees. These organisms are called intermediate species. Eventually, over 150 years, the forest will reach its equilibrium point where species composition is no longer changing and resembles the community before the fire. This equilibrium state is referred to as the climax community, which will remain stable until the next disturbance.

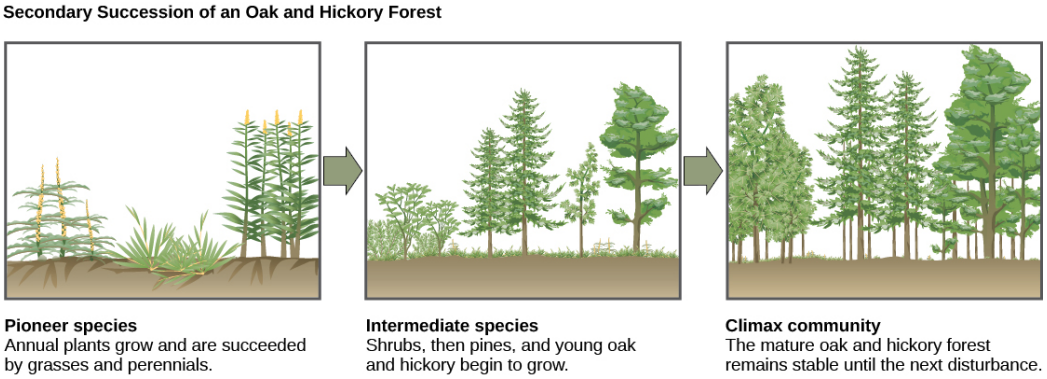


Figure 7.18 Secondary succession is shown in an oak and hickory forest after a forest fire.

8 | ECOLOGICAL RESEARCH

8.1 | Ecosystem Experimentation and Modeling

> "Even if a scientific model, like a car, has only a few years to run before it is discarded, it serves its purpose for getting from one place to another."

David L. Wingate, "Complex Clocks", *Digestive Diseases and Sciences*, 1983, 28:1139

Ecosystems are complex entities, with many components and variables, and thus they present quite a daunting task for ecosystem modelers. Life in an ecosystem is often about competition for limited resources, a characteristic of the mechanism of natural selection. Competition in communities (all living things within specific habitats) is observed both within species and among different species. The resources for which organisms compete include organic material from living or previously living organisms, sunlight, and mineral nutrients, which provide the energy for living processes and the matter to make up organisms' physical structures. Other critical factors influencing community dynamics are the components of its physical and geographic environment: a habitat's latitude, amount of rainfall, topography (elevation), and available species. These are all important environmental variables that determine which organisms can exist within a particular area.

Furthermore, ecosystems are routinely exposed to various disturbances, or changes in the environment that effect their compositions: yearly variations in rainfall and temperature and the slower processes of plant growth, which may take several years. Many of these disturbances are a result of natural processes. For example, when lightning causes a forest fire and destroys part of a forest ecosystem, the ground is eventually populated by grasses, then by bushes and shrubs, and later by mature trees, restoring the forest to its former state. The impact of environmental disturbances caused by human activities is as important as the changes wrought by natural processes. Human agricultural practices, air pollution, acid rain, global deforestation, overfishing, eutrophication, oil spills, and illegal dumping on land and into the ocean are all issues of concern to conservationists.

Research into Ecosystem Dynamics: Ecosystem Experimentation and Modeling

The study of the changes in ecosystem structure caused by changes in the environment (disturbances) or by internal forces is called **ecosystem dynamics**. Ecosystems are characterized using a variety of research methodologies. Some ecologists study ecosystems using controlled experimental systems, while some study entire ecosystems in their natural state, and others use both approaches.

A holistic ecosystem model attempts to quantify the composition, interaction, and dynamics of entire ecosystems; it is the most representative of the ecosystem in its natural state. A food web is an example of a holistic ecosystem model. However, this type of study is limited by time and expense, as well as the fact that it is neither feasible nor ethical to do experiments on large natural ecosystems. To quantify all different species in an ecosystem and the dynamics in their habitat is difficult, especially when studying large habitats such as the Amazon Rainforest, which covers 1.4 billion acres (5.5 million km^2) of the Earth's surface.

For these reasons, scientists study ecosystems under more controlled conditions. Experimental systems usually involve either partitioning a part of a natural ecosystem that can be used for experiments, termed a mesocosm, or by re-creating an ecosystem entirely in an indoor or outdoor laboratory environment, which is referred to as a microcosm. A major limitation to these approaches is that removing individual organisms from their natural ecosystem or altering a natural ecosystem through partitioning may change the dynamics of the ecosystem. These changes are often due to differences in species

numbers and diversity and also to environment alterations caused by partitioning (**mesocosm**) or re-creating (**microcosm**) the natural habitat. Thus, these types of experiments are not totally predictive of changes that would occur in the ecosystem from which they were gathered.

As both of these approaches have their limitations, some ecologists suggest that results from these experimental systems should be used only in conjunction with holistic ecosystem studies to obtain the most representative data about ecosystem structure, function, and dynamics.

Scientists use the data generated by these experimental studies to develop ecosystem models that demonstrate the structure and dynamics of ecosystems. Three basic types of ecosystem modeling are routinely used in research and ecosystem management: a conceptual model, an analytical model, and a simulation model. A conceptual model is an ecosystem model that consists of flow charts to show interactions of different compartments of the living and nonliving components of the ecosystem. A conceptual model describes ecosystem structure and dynamics and shows how environmental disturbances affect the ecosystem; however, its ability to predict the effects of these disturbances is limited. Analytical and simulation models, in contrast, are mathematical methods of describing ecosystems that are indeed capable of predicting the effects of potential environmental changes without direct experimentation, although with some limitations as to accuracy. An analytical model is an ecosystem model that is created using simple mathematical formulas to predict the effects of environmental disturbances on ecosystem structure and dynamics. A simulation model is an ecosystem model that is created using complex computer algorithms to holistically model ecosystems and to predict the effects of environmental disturbances on ecosystem structure and dynamics. Ideally, these models are accurate enough to determine which components of the ecosystem are particularly sensitive to disturbances, and they can serve as a guide to ecosystem managers (such as conservation ecologists or fisheries biologists) in the practical maintenance of ecosystem health.

Conceptual Models

Conceptual models are useful for describing ecosystem structure and dynamics and for demonstrating the relationships between different organisms in a community and their environment. Conceptual models are usually depicted graphically as flow charts. The organisms and their resources are grouped into specific compartments with arrows showing the relationship and transfer of energy or nutrients between them.

To model the cycling of mineral nutrients, organic and inorganic nutrients are subdivided into those that are bioavailable (ready to be incorporated into biological macromolecules) and those that are not. For example, in a terrestrial ecosystem near a deposit of coal, carbon will be available to the plants of this ecosystem as carbon dioxide gas in a short-term period, not from the carbon-rich coal itself. However, over a longer period, microorganisms capable of digesting coal will incorporate its carbon or release it as natural gas (methane, CH_4), changing this unavailable organic source into an available one. This conversion is greatly accelerated by the combustion of fossil fuels by humans, which releases large amounts of carbon dioxide into the atmosphere. This is thought to be a major factor in the rise of the atmospheric carbon dioxide levels in the industrial age. The carbon dioxide released from burning fossil fuels is produced faster than photosynthetic organisms can use it. This process is intensified by the reduction of photosynthetic trees because of worldwide deforestation. Most scientists agree that high atmospheric carbon dioxide is a major cause of global climate change.

Analytical and Simulation Models

The major limitation of conceptual models is their inability to predict the consequences of changes in ecosystem species and/or environment. Ecosystems are dynamic entities and subject to a variety of abiotic and biotic disturbances caused by natural forces and/or human activity. Ecosystems altered from their initial equilibrium state can often recover from such disturbances and return to a state of equilibrium. As most ecosystems are subject to periodic disturbances and are often in a state of change, they are usually either moving toward or away from their equilibrium state. There are many of these equilibrium states among the various components of an ecosystem, which affects the ecosystem overall. Furthermore, as humans have the ability to greatly and rapidly alter the species content and habitat of an ecosystem, the need for predictive models that enable understanding of how ecosystems respond to these changes becomes more crucial.

Analytical models often use simple, linear components of ecosystems, such as food chains, and are known to be complex mathematically; therefore, they require a significant amount of mathematical knowledge and expertise. Although analytical models have great potential, their simplification of complex ecosystems is thought to limit their accuracy. Simulation models that use computer programs are better able to deal with the complexities of ecosystem structure.

A recent development in simulation modeling uses supercomputers to create and run individual-based simulations, which accounts for the behavior of individual organisms and their effects on the ecosystem as a whole. These simulations are considered to be the most accurate and predictive of the complex responses of ecosystems to disturbances.

8.2 | Nitrogen and Phosphorus Cycles

> "As mouths multiply, food resources dwindle. Land is a limited quantity, and the land that will grow wheat is absolutely dependent on difficult and capricious natural phenomena... I have to point the way out of this colossal dilemma. It is the chemist who must come to the rescue of the threatened communities. It is through the laboratory that starvation may ultimately be turned into plenty... The fixation of atmospheric nitrogen is one of the great discoveries, awaiting the genius of chemists."

William Crookes, *Chemical News*. 1898, 78:125.

The "difficult and capricious natural phenomena" of nitrogen fixation is just one part of the biogeochemical cycle of this nutrient, but it is, as Crookes noted, an extremely important part. Chemists have succeeded in the quest to develop reactions that can fix atmospheric nitrogen into forms that can be used by living organisms, but it should be noted that humble microorganisms did this long before boastful chemists thought about it. Industrial processes for nitrogen fixation now produce as much usable nitrogen as all of the biological nitrogen-fixers on the planet. This chemical success, however, does come with some undesirable consequences, as you will learn below.

Energy flows directionally through ecosystems, entering as sunlight (or inorganic molecules for chemoautotrophs) and leaving as heat during the many transfers between trophic levels. However, the matter that makes up living organisms is conserved and recycled. The six most common elements associated with organic molecules—carbon, nitrogen, hydrogen, oxygen, phosphorus, and sulfur—take a variety of chemical forms and may exist for long periods in the atmosphere, on land, in water, or beneath the Earth's surface. Geologic processes, such as weathering, erosion, water drainage, and the subduction of the continental plates, all play a role in this recycling of materials. Because geology and chemistry have major roles in the study of this process, the recycling of inorganic matter between living organisms and their environment is called a **biogeochemical cycle**.

The Nitrogen Cycle

Nitrogen is an essential nutrient for living processes; it is a major component of proteins and nucleic acids. Proteins are important biological molecules because all cellular activities are driven by proteins. Nucleic acids are the building blocks of DNA (hereditary material). Nitrogen is often the limiting nutrient (necessary for growth) on terrestrial ecosystems **Figure 8.1**.

Getting nitrogen into the living world is difficult. Plants and phytoplankton are not equipped to incorporate nitrogen from the atmosphere (which exists as tightly bonded, triple covalent N_2) even though this molecule comprises approximately 78 percent of the atmosphere. Nitrogen enters the living world via free-living and symbiotic bacteria, which incorporate nitrogen into their macromolecules through nitrogen fixation (conversion of N_2). Cyanobacteria live in most aquatic ecosystems where sunlight is present; they play a key role in nitrogen fixation. Cyanobacteria are able to use inorganic sources of nitrogen to "fix" nitrogen. *Rhizobium* bacteria live symbiotically in the root nodules of legumes (such as peas, beans, and peanuts) and provide them with the organic nitrogen they need. Free-living bacteria, such as *Azotobacter*, are also important nitrogen fixers. In addition to natural nitrogen fixation by microbes, humans industrially fix nitrogen to produce artificial fertilizers.

Organic nitrogen is especially important to the study of ecosystem dynamics since many ecosystem processes, such as primary production and decomposition, are limited by the available supply of nitrogen. As shown in **Figure 8.1**, the nitrogen that enters living systems by nitrogen fixation is successively converted from organic nitrogen back into nitrogen gas by bacteria. This process occurs in three steps in terrestrial systems: ammonification, nitrification, and denitrification. First, the ammonification process converts nitrogenous waste from living animals or from the remains of dead animals into ammonium (NH_4^+) by certain bacteria and fungi. Second, the ammonium is converted to nitrites (NO_2^-) by nitrifying

bacteria, such as *Nitrosomonas*, through nitrification. Subsequently, nitrites are converted to nitrates (NO_3^-) by similar organisms. Third, the process of denitrification occurs, whereby bacteria, such as *Pseudomonas* and *Clostridium*, convert the nitrates into nitrogen gas, allowing it to re-enter the atmosphere.

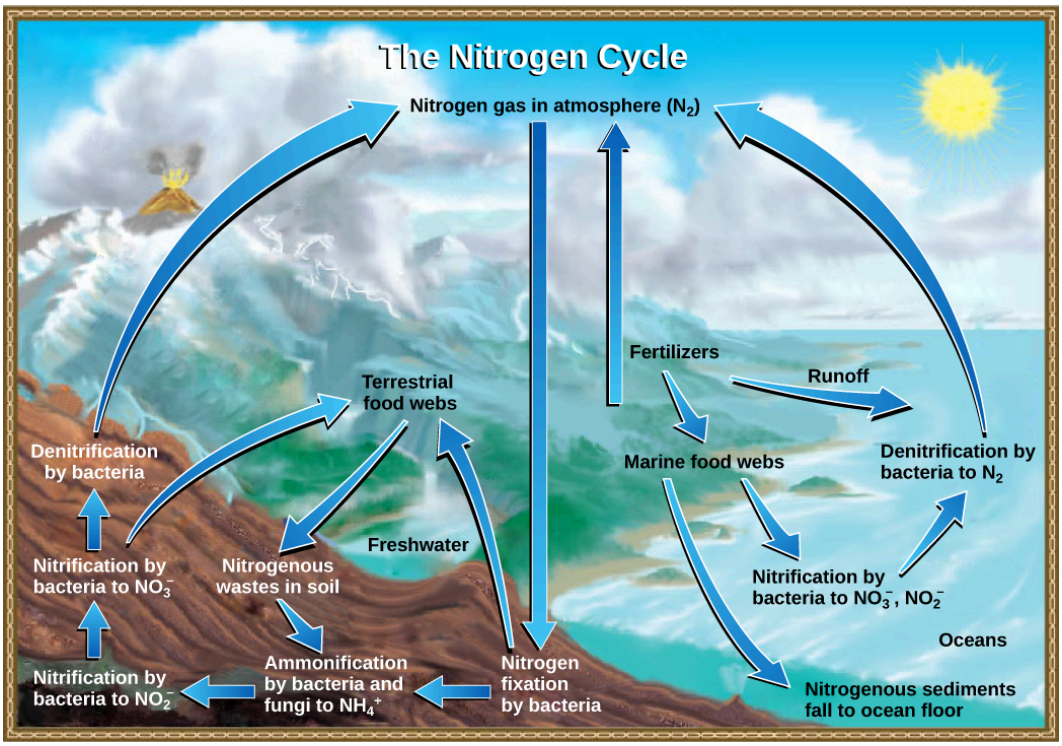


Figure 8.1 Nitrogen enters the living world from the atmosphere via nitrogen-fixing bacteria. This nitrogen and nitrogenous waste from animals is then processed back into gaseous nitrogen by soil bacteria, which also supply terrestrial food webs with the organic nitrogen they need. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

Human activity can release nitrogen into the environment by two primary means: the combustion of fossil fuels, which releases different nitrogen oxides, and by the use of artificial fertilizers in agriculture, which are then washed into lakes, streams, and rivers by surface runoff. Atmospheric nitrogen is associated with several effects on Earth's ecosystems including the production of acid rain (as nitric acid, HNO_3) and greenhouse gas (as nitrous oxide, N_2O) potentially causing climate change. A major effect from fertilizer runoff is saltwater and freshwater **eutrophication**, a process whereby nutrient runoff causes the excess growth of microorganisms, depleting dissolved oxygen levels and killing ecosystem fauna.

A similar process occurs in the marine nitrogen cycle, where the ammonification, nitrification, and denitrification processes are performed by marine bacteria. Some of this nitrogen falls to the ocean floor as sediment, which can then be moved to land in geologic time by uplift of the Earth's surface and thereby incorporated into terrestrial rock. Although the movement of nitrogen from rock directly into living systems has been traditionally seen as insignificant compared with nitrogen fixed from the atmosphere, a recent study showed that this process may indeed be significant and should be included in any study of the global nitrogen cycle.[1]

The Phosphorus Cycle

Phosphorus is another essential nutrient for living processes; it is a major component of nucleic acids, phospholipids, and, as calcium phosphate, makes up the supportive components of our bones. Phosphorus is often the limiting nutrient (necessary for growth) in aquatic ecosystems (**Figure 8.2**).

Phosphorus occurs in nature as the phosphate ion (PO_4^{3-}). In addition to phosphate runoff as a result of human activity (mined to make artificial fertilizers), natural surface runoff occurs when it is leached from phosphate-containing rock by weathering, thus sending phosphates into rivers, lakes, and the ocean. This rock has its origins in the ocean. Phosphate-containing ocean sediments form primarily from the bodies of ocean organisms and from their excretions. However, in remote regions, volcanic ash, aerosols, and mineral dust may also be significant phosphate sources. This sediment then is

1. Scott L. Morford, Benjamin Z. Houlton, and Randy A. Dahlgren, "Increased Forest Ecosystem Carbon and Nitrogen Storage from Nitrogen Rich Bedrock," *Nature* 477, no. 7362 (2011): 78–81.

moved to land over geologic time by the uplifting of areas of the Earth's surface.

Phosphorus is also reciprocally exchanged between phosphate dissolved in the ocean and marine ecosystems. The movement of phosphate from the ocean to the land and through the soil is extremely slow, with the average phosphate ion having an oceanic residence time between 20,000 and 100,000 years.

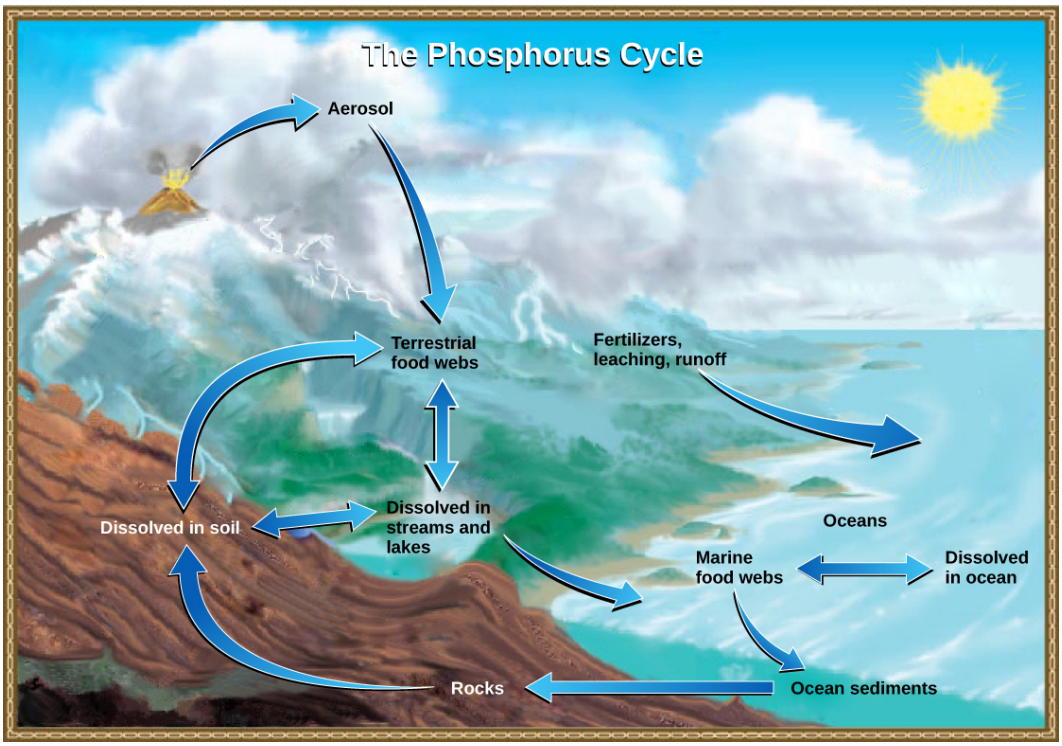


Figure 8.2 In nature, phosphorus exists as the phosphate ion (PO_4^{3-}). Weathering of rocks and volcanic activity releases phosphate into the soil, water, and air, where it becomes available to terrestrial food webs. Phosphate enters the oceans via surface runoff, groundwater flow, and river flow. Phosphate dissolved in ocean water cycles into marine food webs. Some phosphate from the marine food webs falls to the ocean floor, where it forms sediment. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

Excess phosphorus and nitrogen that enters these ecosystems from fertilizer runoff and from sewage causes excessive growth of microorganisms and depletes the dissolved oxygen, which leads to the death of many ecosystem fauna, such as shellfish and finfish. This process is responsible for dead zones in lakes and at the mouths of many major rivers (**Figure 8.2**).

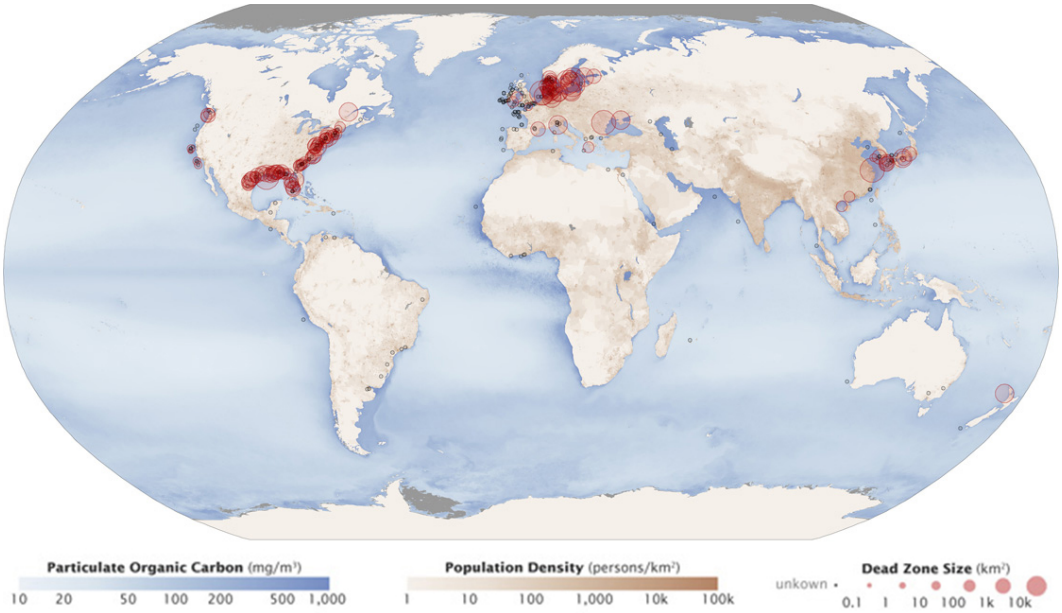


Figure 8.3 Dead zones occur when phosphorus and nitrogen from fertilizers cause excessive growth of microorganisms, which depletes oxygen and kills fauna. Worldwide, large dead zones are found in coastal areas of high population density. (credit: NASA Earth Observatory)

A **dead zone** is an area within a freshwater or marine ecosystem where large areas are depleted of their normal flora and fauna; these zones can be caused by eutrophication, oil spills, dumping of toxic chemicals, and other human activities. The number of dead zones has been increasing for several years, and more than 400 of these zones were present as of 2008. One of the worst dead zones is off the coast of the United States in the Gulf of Mexico, where fertilizer runoff from the Mississippi River basin has created a dead zone of over 8463 square miles. Phosphate and nitrate runoff from fertilizers also negatively affect several lake and bay ecosystems including the Chesapeake Bay in the eastern United States.

everyday CONNECTION

Chesapeake Bay

(a) (b)

Figure 8.4 This (a) satellite image shows the Chesapeake Bay, an ecosystem affected by phosphate and nitrate runoff. A (b) member of the Army Corps of Engineers holds a clump of oysters being used as a part of the oyster restoration effort in the bay. (credit a: modification of work by NASA/MODIS; credit b: modification of work by U.S. Army)

The Chesapeake Bay has long been valued as one of the most scenic areas on Earth; it is now in distress and is recognized as a declining ecosystem. In the 1970s, the Chesapeake Bay was one of the first ecosystems to have identified dead zones, which continue to kill many fish and bottom-dwelling species, such as clams, oysters, and worms. Several species have declined in the Chesapeake Bay due to surface water runoff containing excess nutrients from artificial fertilizer used on land. The source of the fertilizers (with high nitrogen and phosphate content) is not limited to agricultural practices. There are many nearby urban areas and more than 150 rivers and streams empty into the bay that are carrying fertilizer runoff from lawns and gardens. Thus, the decline of the Chesapeake Bay is a complex issue and requires the cooperation of industry, agriculture, and everyday homeowners.

Of particular interest to conservationists is the oyster population; it is estimated that more than 200,000 acres of oyster reefs existed in the bay in the 1700s, but that number has now declined to only 36,000 acres. Oyster harvesting was once a major industry for Chesapeake Bay, but it declined 88 percent between 1982 and 2007. This decline was due not only to fertilizer runoff and dead zones but also to overharvesting. Oysters require a certain minimum population density because they must be in close proximity to reproduce. Human activity has altered the oyster population and locations, greatly disrupting the ecosystem.

The restoration of the oyster population in the Chesapeake Bay has been ongoing for several years with mixed success. Not only do many people find oysters good to eat, but they also clean up the bay. Oysters are filter feeders, and as they eat, they clean the water around them. In the 1700s, it was estimated that it took only a few days for the oyster population to filter the entire volume of the bay. Today, with changed water conditions, it is estimated that the present population would take nearly a year to do the same job.

Restoration efforts have been ongoing for several years by non-profit organizations, such as the Chesapeake Bay Foundation. The restoration goal is to find a way to increase population density so the oysters can reproduce more efficiently. Many disease-resistant varieties (developed at the Virginia Institute of Marine Science for the College of William and Mary) are now available and have been used in the construction of experimental oyster reefs. Efforts to clean and restore the bay by Virginia and Delaware have been hampered because much of the pollution entering the bay comes from other states, which stresses the need for inter-state cooperation to gain successful restoration.

The new, hearty oyster strains have also spawned a new and economically viable industry—oyster

aquaculture—which not only supplies oysters for food and profit, but also has the added benefit of cleaning the bay.

8.3 | Freshwater Biomes

Introduction

> "For as the element of water lies in the middle of the globe, so, the branches run out from the root in its circuit on all sides towards the plains and towards the light. From this root very many branches are born."

Paracelsus, "The Philosophy of the Generation of the Elements", Book the Fourth, Text II. In *The Hermetic and Alchemical Writings of Aureolus Phillipus Theophrastus Bombast, of Hohenheim, called Paracelsus the Great*, translated by A.E. Waite (1894), 1:232.

The alchemical thinking of Paracelsus might seem unscientific today, but his insights about the central nature of water are still viable. Freshwater biomes are among the most important on the planet in terms of species diversity and ecosystem services. Abiotic and biotic (including human) impacts on these biomes are among the most important factors in influencing these functions, and their role in the various biogeochemical cycles cannot be overstated.

Abiotic Factors Influencing Aquatic Biomes

Aquatic biomes are influenced by a series of abiotic factors associated with water, and these factors include the amount of light, stratification due to temperature, and the thermal properties of water. Another abiotic factor is nutrients, review the following information about freshwater biomes and begin to think about how human disturbances can affect freshwater ecosystems.

Freshwater Biomes

Freshwater biomes include lakes and ponds (standing water) as well as rivers and streams (flowing water). They also include wetlands, which will be discussed later. Humans rely on freshwater biomes to provide aquatic resources for drinking water, crop irrigation, sanitation, and industry. These various roles and human benefits are referred to as ecosystem services. Lakes and ponds are found in terrestrial landscapes and are, therefore, connected with abiotic and biotic factors influencing these terrestrial biomes.

Lakes and Ponds

Lakes and ponds can range in area from a few square meters to thousands of square kilometers. Temperature is an important abiotic factor affecting living things found in lakes and ponds. In the summer, thermal stratification of lakes and ponds occurs when the upper layer of water is warmed by the sun and does not mix with deeper, cooler water. Light can penetrate within the photic zone of the lake or pond. Phytoplankton (algae and cyanobacteria) are found here and carry out photosynthesis, providing the base of the food web of lakes and ponds. Zooplankton, such as rotifers and small crustaceans, consume these phytoplankton. At the bottom of lakes and ponds, bacteria in the aphotic zone break down dead organisms that sink to the bottom.

Nitrogen and phosphorus are important limiting nutrients in lakes and ponds. Because of this, they are determining factors in the amount of phytoplankton growth in lakes and ponds. When there is a large input of nitrogen and phosphorus (from sewage and runoff from fertilized lawns and farms, for example), the growth of algae skyrockets, resulting in a large accumulation of algae called an algal bloom. Algal blooms can become so extensive that they reduce light penetration in water. As a result, the lake or pond becomes aphotic and photosynthetic plants cannot survive. When the algae die and decompose, severe oxygen depletion of the water occurs. Fishes and other organisms that require oxygen are then more likely to die, and resulting dead zones are found across the globe. Lake Erie and the Gulf of Mexico represent freshwater and marine habitats where phosphorus control and storm water runoff pose significant environmental challenges.

Rivers and Streams

Rivers and streams are continuously moving bodies of water that carry large amounts of water from the source, or headwater, to a lake or ocean. The largest rivers include the Nile River in Africa, the Amazon River in South America, and the Mississippi River in North America.

Abiotic features of rivers and streams vary along the length of the river or stream. Streams begin at a point of origin referred to as source water. The source water is usually cold, low in nutrients, and clear. The channel (the width of the river or stream) is narrower than at any other place along the length of the river or stream. Because of this, the current is often faster here than at any other point of the river or stream.

The fast-moving water results in minimal silt accumulation at the bottom of the river or stream; therefore, the water is clear. Photosynthesis here is mostly attributed to algae that are growing on rocks; the swift current inhibits the growth of phytoplankton. An additional input of energy can come from leaves or other organic material that falls into the river or stream from trees and other plants that border the water. When the leaves decompose, the organic material and nutrients in the leaves are returned to the water. Plants and animals have adapted to this fast-moving water. For instance, leeches (phylum Annelida) have elongated bodies and suckers on both ends. These suckers attach to the substrate, keeping the leech anchored in place. Freshwater trout species (phylum Chordata) are an important predator in these fast-moving rivers and streams.

As the river or stream flows away from the source, the width of the channel gradually widens and the current slows. This slow-moving water, caused by the gradient decrease and the volume increase as tributaries unite, has more sedimentation. Phytoplankton can also be suspended in slow-moving water. Therefore, the water will not be as clear as it is near the source. The water is also warmer. Worms (phylum Annelida) and insects (phylum Arthropoda) can be found burrowing into the mud. The higher order predator vertebrates (phylum Chordata) include waterfowl, frogs, and fishes. These predators must find food in these slow moving, sometimes murky, waters and, unlike the trout in the waters at the source, these vertebrates may not be able to use vision as their primary sense to find food. Instead, they are more likely to use taste or chemical cues to find prey.

Wetlands

Wetlands are environments in which the soil is either permanently or periodically saturated with water. Wetlands are different from lakes because wetlands are shallow bodies of water whereas lakes vary in depth. Emergent vegetation consists of wetland plants that are rooted in the soil but have portions of leaves, stems, and flowers extending above the water's surface. There are several types of wetlands including marshes, swamps, bogs, mudflats, and salt marshes. The three shared characteristics among these types—what makes them wetlands—are their hydrology, hydrophytic vegetation, and hydric soils.

Freshwater marshes and swamps are characterized by slow and steady water flow. Bogs develop in depressions where water flow is low or nonexistent. Bogs usually occur in areas where there is a clay bottom with poor percolation. Percolation is the movement of water through the pores in the soil or rocks. The water found in a bog is stagnant and oxygen depleted because the oxygen that is used during the decomposition of organic matter is not replaced. As the oxygen in the water is depleted, decomposition slows. This leads to organic acids and other acids building up and lowering the pH of the water. At a lower pH, nitrogen becomes unavailable to plants. This creates a challenge for plants because nitrogen is an important limiting resource. Some types of bog plants (such as sundews, pitcher plants, and Venus flytraps) capture insects and extract the nitrogen from their bodies. Bogs have low net primary productivity because the water found in bogs has low levels of nitrogen and oxygen.

8.4 | Population Growth Curves

Introduction

> "Population, when unchecked, increases in a geometrical ratio. Subsistence increases only in an arithmetical ratio. A slight acquaintance with numbers will show the immensity of the first power in comparison with the second."

Thomas Malthus, *An Essay on the Principle of Populations*, 1798

Malthus recognized the fact that there is a connection between resources and population growth, and that one of these (population growth) can increase at a greater rate than the other. Modern population ecologists make use of a variety of methods to model population dynamics mathematically. These more precise models can then be used to accurately describe changes occurring in a population and better predict future changes. Use the following information to make sure that you have at least a "slight acquaintance" with the mathematical principles that are used to describe population growth.

Exponential Growth

Charles Darwin, in his theory of natural selection, was greatly influenced by the English clergyman Thomas Malthus. Malthus published a book in 1798 stating that populations with unlimited natural resources grow very rapidly. This accelerating pattern of increasing population size is called **exponential growth**.

The best example of exponential growth is seen in bacteria. Some species of Bacteria can undergo cell division about every hour. If 1000 bacteria are placed in a large flask with an unlimited supply of nutrients (so the nutrients will not become depleted), after an hour, there is one round of division and each organism divides, resulting in 2000 organisms—an increase of 1000. In another hour, each of the 2000 organisms will double, producing 4000, an increase of 2000 organisms. After the third hour, there should be 8000 bacteria in the flask, an increase of 4000 organisms. The important concept of exponential growth is that the **population growth rate**—the number of organisms added in each reproductive generation—is accelerating; that is, it is increasing at a greater and greater rate. After 1 day and 24 of these cycles, the population would have increased from 1000 to more than 16 billion. When the population size, *N*, is plotted over time, a **J-shaped growth curve** is produced (**Figure 8.5**).

Exponential growth is common when population organisms have unlimited resources. The growth of that population can be calculated using the equation below. For further explanation of this equation please go to **Population Growth** .

$$G = rN$$

The value *"r"* can be positive, meaning the population is increasing in size; or negative, meaning the population is decreasing in size; or zero, where the population's size is unchanging, a condition known as zero population growth. A further refinement of the formula recognizes that different species have inherent differences in their per capita growth rates (*r*), even under ideal conditions. Obviously, a bacterium can reproduce more rapidly and have a higher per capita growth rate than a human. The maximal growth rate for a species is its biotic potential, or r_{max}, thus changing the equation to:

$$G = r_{\max} N$$

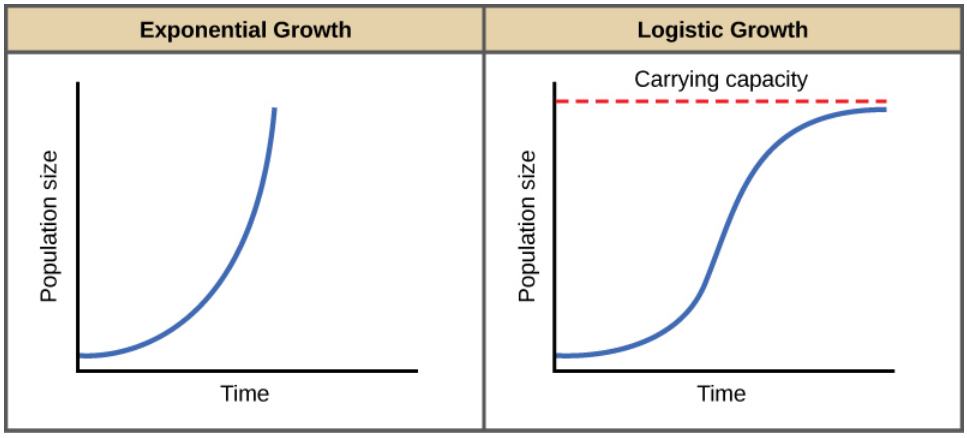


Figure 8.5 When resources are unlimited, populations exhibit exponential growth, resulting in a J-shaped curve. When resources are limited, populations exhibit logistic growth. In logistic growth, population expansion decreases as resources become scarce, and it levels off when the carrying capacity of the environment is reached, resulting in an S-shaped curve.

Logistic Growth

Exponential growth is possible only when infinite natural resources are available; this is not the case in the real world. Charles Darwin recognized this fact in his description of the "struggle for existence," which states that individuals will compete (with members of their own or other species) for limited resources. The successful ones will survive to pass on their own characteristics and traits (which we know now are transferred by genes) to the next generation at a greater rate (natural selection). To model the reality of limited resources, population ecologists developed the **logistic growth** model.

Carrying Capacity and the Logistic Model

In the real world, with its limited resources, exponential growth cannot continue indefinitely. Exponential growth may occur in environments where there are few individuals and plentiful resources, but when the number of individuals gets large enough, resources will be depleted, slowing the growth rate. Eventually, the growth rate will plateau or level off (**Figure 8.5**). This population size, which represents the maximum population size that a particular environment can support, is called the **carrying capacity, or *K***.

The formula we use to calculate logistic growth adds the carrying capacity as a moderating force in the growth rate. The expression "$K - N$" is indicative of how many individuals may be added to a population at a given stage, and "$K - N$" divided by "K" is the fraction of the carrying capacity available for further growth. Thus, the exponential growth model is restricted by this factor to generate the logistic growth equation:

$$G = r_{\max} N \frac{(K - N)}{K}$$

Notice that when N is very small, $(K\text{-}N)/K$ becomes close to K/K or 1, and the right side of the equation reduces to $r_{max}N$, which means the population is growing exponentially and is not influenced by carrying capacity. On the other hand, when N is large, $(K\text{-}N)/K$ come close to zero, which means that population growth will be slowed greatly or even stopped. Thus, population growth is greatly slowed in large populations by the carrying capacity K. This model also allows for the population of a negative population growth, or a population decline. This occurs when the number of individuals in the population exceeds the carrying capacity (because the value of (K-N)/K is negative).

A graph of this equation yields an **S-shaped curve** (**Figure 8.5**), and it is a more realistic model of population growth than exponential growth. There are three different sections to an S-shaped curve. Initially, growth is exponential because there are few individuals and ample resources available. Then, as resources begin to become limited, the growth rate decreases. Finally, growth levels off at the carrying capacity of the environment, with little change in population size over time.

8.5 | Introduction to Water Pollution

The Water Pollution Crisis

Introduction

- Pollution, pollution,
- You can use the latest toothpaste,
- And then rinse your mouth with industrial waste.

Tom Lehrer, lyrics from his 1965 song *Pollution*

The effects humans have on the water cycle and fresh water supply described one aspect of the global water crisis, specifically the water shortages that afflict many arid and densely populated areas. The global water crisis also involves water pollution, because to be useful for drinking and irrigation, water must not be polluted beyond certain thresholds. According to the World Health Organization, in 2008 approximately 880 million people in the world (or 13% of world population) did not have access to improved (safe) drinking water (**World Health Statistics, 2010**) (**Figure 8.6**). At the same time, about 2.6 billion people (or 40% of world population) lived without improved sanitation (**Figure 8.7**), which is defined as having access to a public sewage system, septic tank, or even a simple pit latrine. Each year approximately 1.7 million people die from diarrheal diseases associated with unsafe drinking water, inadequate sanitation, and poor hygiene, e.g., hand washing with soap. Almost all of these deaths are in developing countries, and around 90% of them occur among children under the age of 5 (see Figure **Figure 8.8**). Compounding the water crisis is the issue of social justice; poor people more commonly lack clean water and sanitation than wealthy people in similar areas. Globally, improving water, sanitation, and hygiene could prevent up to 9% of all disease and 6% of all deaths. In addition to the global waterborne disease crisis, chemical pollution from agriculture, industry, cities, and mining threatens global water quality. Some chemical pollutants have serious and well-known health effects; however, many others have poorly known long-term health effects. In the U.S. currently more than 40,000 water bodies fit the definition of "impaired" set by EPA (See Figure **Figure 8.9**), which means they could neither support a healthy ecosystem nor meet water quality standards. In Gallup public polls conducted over the past decade Americans consistently put water pollution and water supply as the top environmental concerns over issues such as air pollution, deforestation, species extinction, and global warming.

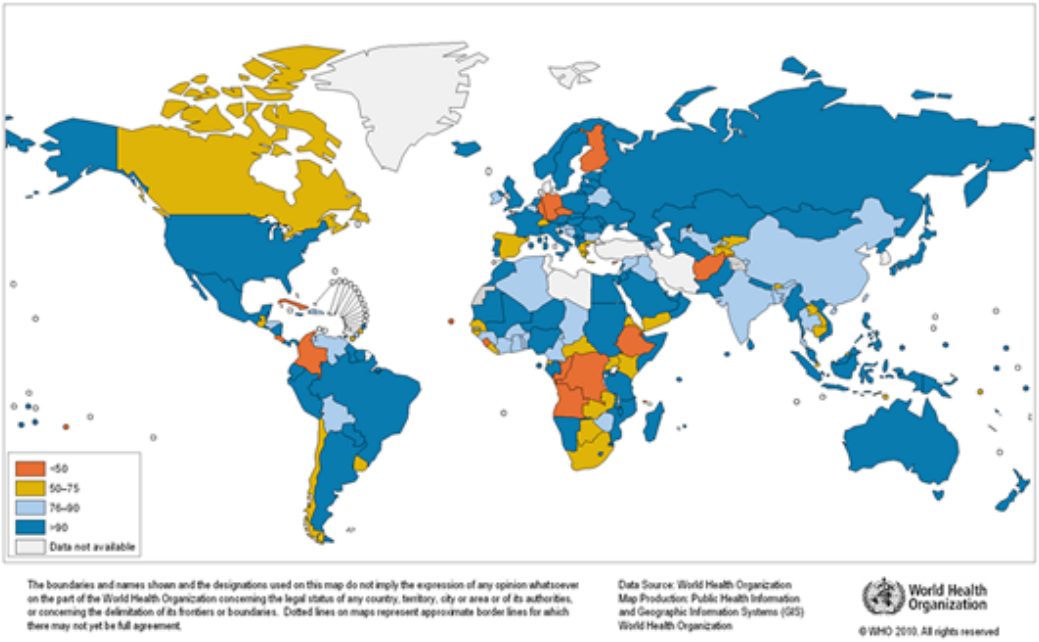


Figure 8.6 Proportion of Population by Country Using Improved Drinking Water Sources in 2008 Improved drinking water sources, e.g., household connections, public standpipes, boreholes, protected dug wells and springs, and rainwater collections, are defined as those more likely to provide safe water than unimproved water sources, e.g., unprotected wells and springs, vendor-provided water, bottled water (unless water for other uses is available from an improved source), and tanker truck-provided water. *Source:* ***World Health Organization (http://gamapserver.who.int/mapLibrary/Files/Maps/phe_water_08.png)***

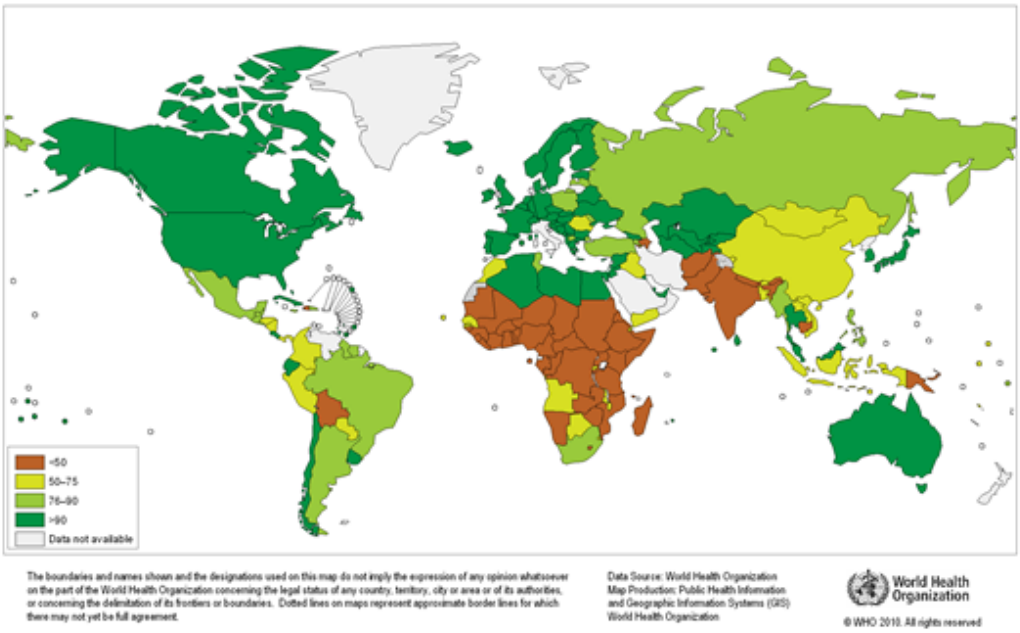


Figure 8.7 Proportion of Population by Country Using Improved Sanitation Facilities in 2008 Improved sanitation facilities, e.g., connection to public sewers or septic systems, pour-flush latrines, pit latrines, and ventilated improved pit latrines, are defined as those more likely to be sanitary than unimproved facilities, e.g., bucket latrines, public latrines, and open pit latrines. *Source:* ***World Health Organization (http://gamapserver.who.int/mapLibrary/Files/Maps/MDG7_sanitation_08.png)***

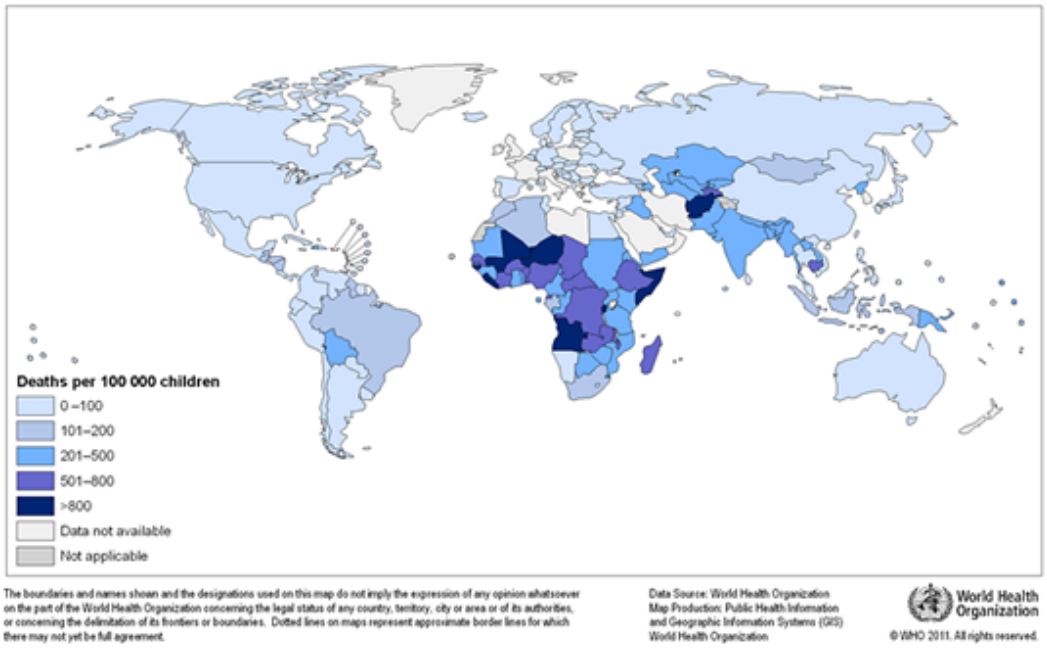


Figure 8.8 Deaths by Country from Diarrhea Caused by Unsafe Water, Unimproved Sanitation, and Poor Hygiene in Children Less than 5 Years Old, 2004*Source:* ***World Health Organization (http://gamapserver.who.int/mapLibrary/Files/Maps/Global_wsh_death_under5_2004.png)***

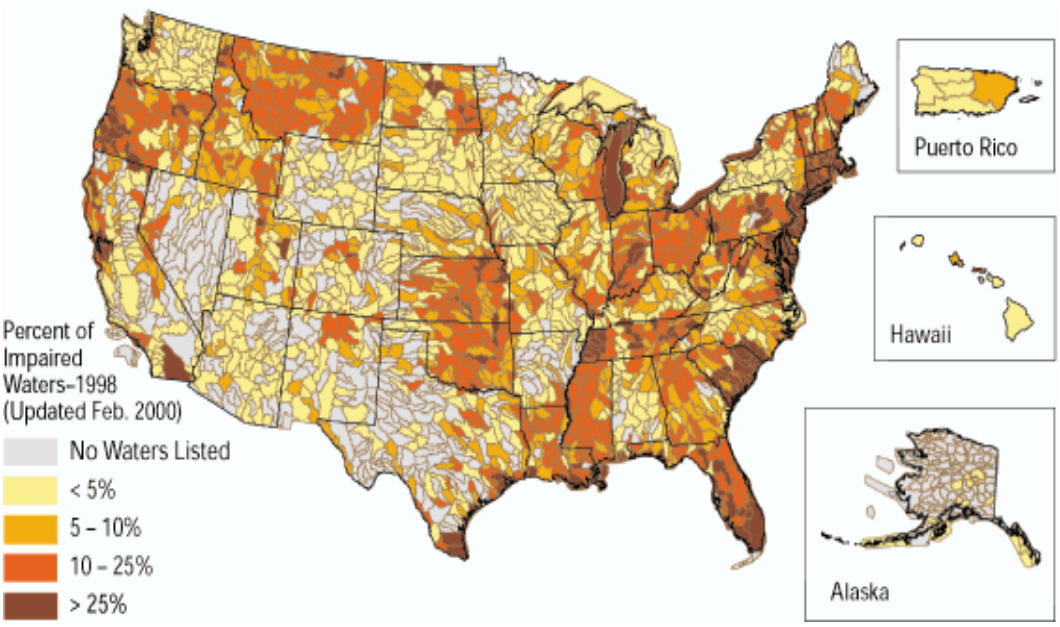


Figure 8.9 Percentage of Impaired Water Bodies in a Watershed by State in USA Based on US EPA Data in 2000 Map of watersheds containing impaired water bodies from the U.S. Environmental Protection Agency's 1998 list of impaired waters *Source:* ***U.S. Geological Survey (http://pubs.usgs.gov/fs/FS-130-01/)***

Water Chemistry Overview

Compared to other molecules of similar molecular weight, water (H_2O) has unique physical properties including high values for melting and boiling point, surface tension (water's cohesion, or "stickiness"), and capacity to dissolve soluble minerals, i.e., act as a solvent. These properties are related to its asymmetrical structure and *polar nature*, which means it is electrically neutral overall but it has a net positive charge on the side with the two hydrogen atoms and a net negative charge on the oxygen side (**Figure 8.10**). This separation of the electrical charge within a water molecule results in *hydrogen bonds* with other water molecules, mineral surfaces (hydrogen bonding produces the water films on minerals in the unsaturated zone

of the subsurface), and *dissolved ions* (atoms with a negative or positive charge). Many minerals and pollutants dissolve readily in water because water forms *hydration shells* (spheres of loosely coordinated, oriented water molecules) around ions.

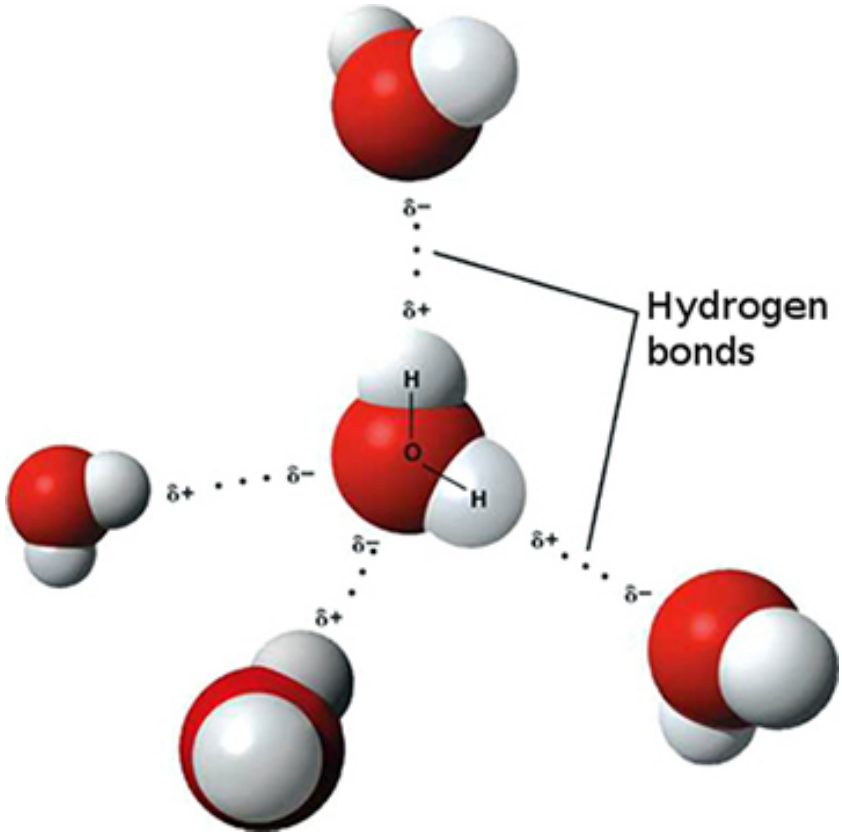


Figure 8.10 Structure of Water, Polar Charge of Water, and Hydrogen Bonds between Water Molecules *Source: Michal Maňas (http://commons.wikimedia.org/wiki/File:3D_model_hydrogen_bonds_in_water.jpg) at Wikimedia Commons*

Any natural water contains dissolved chemicals; some of these are important human nutrients, while others can be harmful to human health. The abundance of a water pollutant is commonly given in very small concentration units such as parts per million (ppm) or even parts per billion (ppb). An arsenic concentration of 1 ppm means 1 part of arsenic per million parts of water. This is equivalent to one drop of arsenic in 50 liters of water. To give you a different perspective on appreciating small concentration units, converting 1 ppm to length units is 1 cm (0.4 in) in 10 km (6 miles) and converting 1 ppm to time units is 30 seconds in a year. Total dissolved solids (TDS) represent the total amount of dissolved material in water. Average TDS (salinity) values for rainwater, river water, and seawater are about 4 ppm, 120 ppm, and 35,000 ppm. The most important processes that affect the salinity of natural waters are evaporation, which distills nearly pure water and leaves the dissolved ions in the original water, and chemical weathering, which involves mineral dissolution that adds dissolved ions to water. Fresh water is commonly defined as containing less than either 1,000 or 500 ppm TDS, but the US Environmental Protection Agency (EPA) recommends that drinking water not exceed 500 ppm TDS or else it will have an unpleasant salty taste.

Water Pollution Overview

Water pollution is the contamination of water by an excess amount of a substance that can cause harm to human beings and the ecosystem. The level of water pollution depends on the abundance of the pollutant, the ecological impact of the pollutant, and the use of the water. Pollutants are derived from biological, chemical, or physical processes. Although natural processes such as volcanic eruptions or evaporation sometimes can cause water pollution, most pollution is derived from human, land-based activities (see Figure **Figure 8.11**). Water pollutants can move through different water reservoirs, as the water carrying them progresses through stages of the water cycle (see Figure **Figure 8.12**). Water residence time (the average time that a water molecule spends in a water reservoir) is very important to pollution problems because it affects pollution potential. Water in rivers has a relatively short residence time, so pollution usually is there only briefly. Of course, pollution in rivers may simply move to another reservoir, such as the ocean, where it can cause further problems.

Groundwater is typically characterized by slow flow and longer residence time, which can make groundwater pollution particularly problematic. Finally, pollution residence time can be much greater than the water residence time because a pollutant may be taken up for a long time within the ecosystem or absorbed onto sediment.

Figure 8.11 Water Pollution Obvious water pollution in the form of floating debris; invisible water pollutants sometimes can be much more harmful than visible ones. *Source: **Stephen Codrington (http://commons.wikimedia.org/?title=File:Obvious_water_pollution.jpeg)** at Wikimedia Commons*

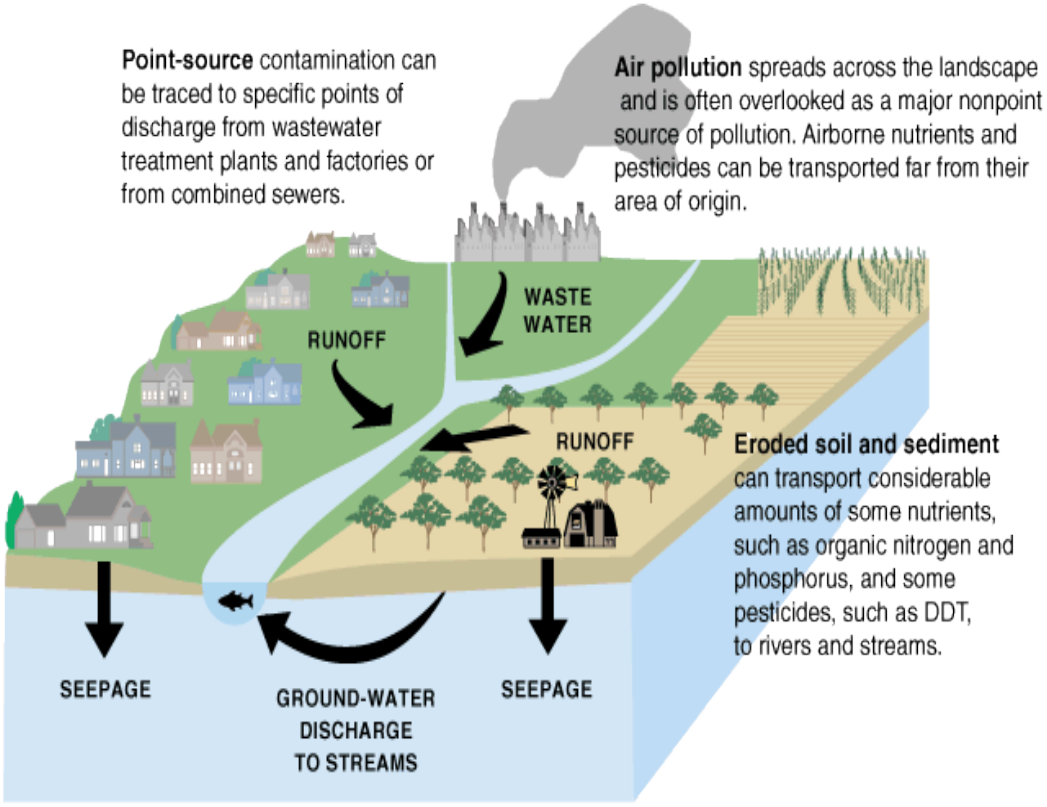


Figure 8.12 Sources of Water Contamination Sources of some water pollutants and movement of pollutants into different water reservoirs of the water cycle. *Source: U.S. Geological Survey (http://ga.water.usgs.gov/edu/waterquality.html)*

Pollutants enter water supplies from point sources, which are readily identifiable and relatively small locations, or nonpoint sources, which are large and more diffuse areas. Point sources of pollution include animal "factory" farms that raise a large number and high density of livestock such as cows, pigs, and chickens (**Figure 8.13**) and discharge pipes from a factories or sewage treatment plants. Combined sewer systems that have a single set of underground pipes to collect both sewage and storm water runoff from streets for wastewater treatment can be major point sources of pollutants. During heavy rain, storm water runoff may exceed sewer capacity, causing it to back up and spilling untreated sewage into surface waters (**Figure 8.14**). Nonpoint sources of pollution include agricultural fields, cities, and abandoned mines. Rainfall runs over the land and through the ground, picking up pollutants such as herbicides, pesticides, and fertilizer from agricultural fields and lawns; oil, antifreeze, car detergent, animal waste, and road salt from urban areas; and acid and toxic elements from abandoned mines. Then, this pollution is carried into surface water bodies and groundwater. Nonpoint source pollution, which is the leading cause of water pollution in the U.S., is usually much more difficult and expensive to control than point source pollution because of its low concentration, multiple sources, and much greater volume of water.

Figure 8.13 A Commercial Meat Chicken Production House This chicken factory farm is a possible major point source of water pollution. *Source: Larry Rana (http://commons.wikimedia.org/wiki/File:Florida_chicken_house.jpg) at Wikimedia Commons*

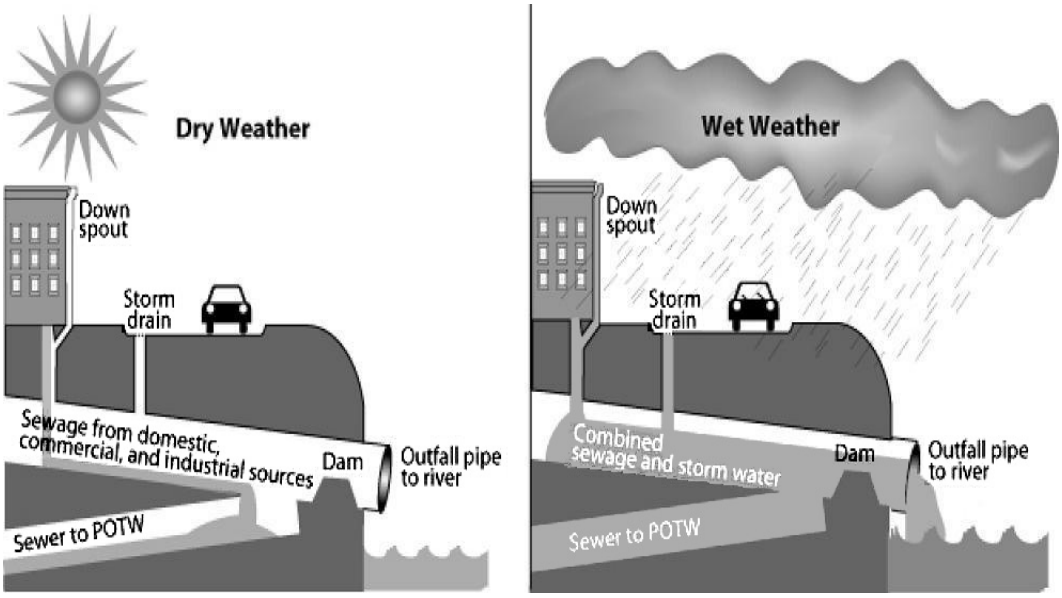


Figure 8.14 Combined Sewer System A combined sewer system is a possible major point source of water pollution during heavy rain due to overflow of untreated sewage. During dry weather (and small storms), all flows are handled by the publicly owned treatment works (POTW). During large storms, the relief structure allows some of the combined stormwater and sewage to be discharged untreated to an adjacent water body. *Source: U.S. Environmental Protection Agency (http://commons.wikimedia.org/wiki/File:CSO_diagram_US_EPA.jpg) at Wikimedia Commons*

Types of Water Pollutants

Oxygen-demanding waste is an extremely important pollutant to ecosystems. Most surface water in contact with the atmosphere has a small amount of dissolved oxygen, which is needed by aquatic organisms for cellular respiration. Bacteria

decompose dead organic matter (chemically represented in a simplified way as CH_2O) and remove dissolved oxygen (O_2) according to the following reaction:

$$CH_2O + O_2 \rightarrow CO_2 + H_2O$$

Too much decaying organic matter in water is a pollutant because it removes oxygen from water, which can kill fish, shellfish, and aquatic insects. The amount of oxygen used by aerobic (in the presence of oxygen) bacterial decomposition of organic matter is called biochemical oxygen demand (BOD). The major source of dead organic matter in most natural waters is sewage; grass and leaves are smaller sources. An unpolluted water body with respect to oxygen is a turbulent river that flows through a natural forest. Turbulence continually brings water in contact with the atmosphere where the O_2 content is restored. The dissolved oxygen content in such a river ranges from 10 to 14 ppm O_2, BOD is low, and clean-water fish, e.g., bass, trout, and perch dominate. A polluted water body with respect to oxygen is a stagnant deep lake in an urban setting with a combined sewer system. This system favors a high input of dead organic carbon from sewage overflows and limited chance for water circulation and contact with the atmosphere. In such a lake, the dissolved O_2 content is ≤5 ppm O_2, BOD is high, and low O_2-tolerant fish, e.g., carp and catfish dominate.

Excessive plant nutrients, particularly nitrogen (N) and phosphorous (P), are pollutants closely related to oxygen-demanding waste. Aquatic plants require about 15 nutrients for growth, most of which are plentiful in water. N and P are called *limiting nutrients*, because they usually are present in water at low concentrations and therefore restrict the total amount of plant growth. This explains why N and P are major ingredients in most fertilizer. High concentrations of N and P from human sources (mostly agricultural and urban runoff including fertilizer, sewage, and P-based detergent) can cause cultural eutrophication, which involves the rapid growth of aquatic plants, particularly algae, called an *algal bloom*. Thick mats of floating and rooted green or sometimes red algae (**Figure 8.15**) create water pollution, damage the ecosystem by clogging fish gills and blocking sunlight, and damage lake aesthetics by making recreation difficult and creating an eyesore. A small percentage of algal species produce toxins that can kill fish, mammals, and birds, and may cause human illness; explosive growths of these algae are called *harmful algal blooms* (**Figure 8.16**). When the prolific algal layer dies, it becomes oxygen-demanding waste, which can create very low O_2 water (<~2 ppm O_2), called hypoxia or dead zone because it causes death to organisms that are unable to leave that environment. An estimated 50% of lakes in North America, Europe, and Asia are negatively impacted by cultural eutrophication. In addition, the size and number of marine hypoxic zones have grown dramatically over the past 50 years (**Figure 8.17**), including a very large dead zone located offshore Louisiana in the Gulf of Mexico. Cultural eutrophication and hypoxia are difficult to combat, because they are caused primarily by nonpoint source pollution, which is difficult to regulate, and N and P, which are difficult to remove from wastewater.

Figure 8.15 Algal Bloom in River in Sichuan, China Algal blooms can present problems for ecosystems and human society. *Source: **Felix Andrews (http://commons.wikimedia.org/wiki/File:River_algae_Sichuan.jpg)** via Wikimedia Commons*

Figure 8.16 Harmful Algal Bloom Harmful algal bloom with deep red color. *Source: Kai Schumann (http://oceanservice.noaa.gov/hazards/hab/) via National Oceanic and Atmospheric Administration*

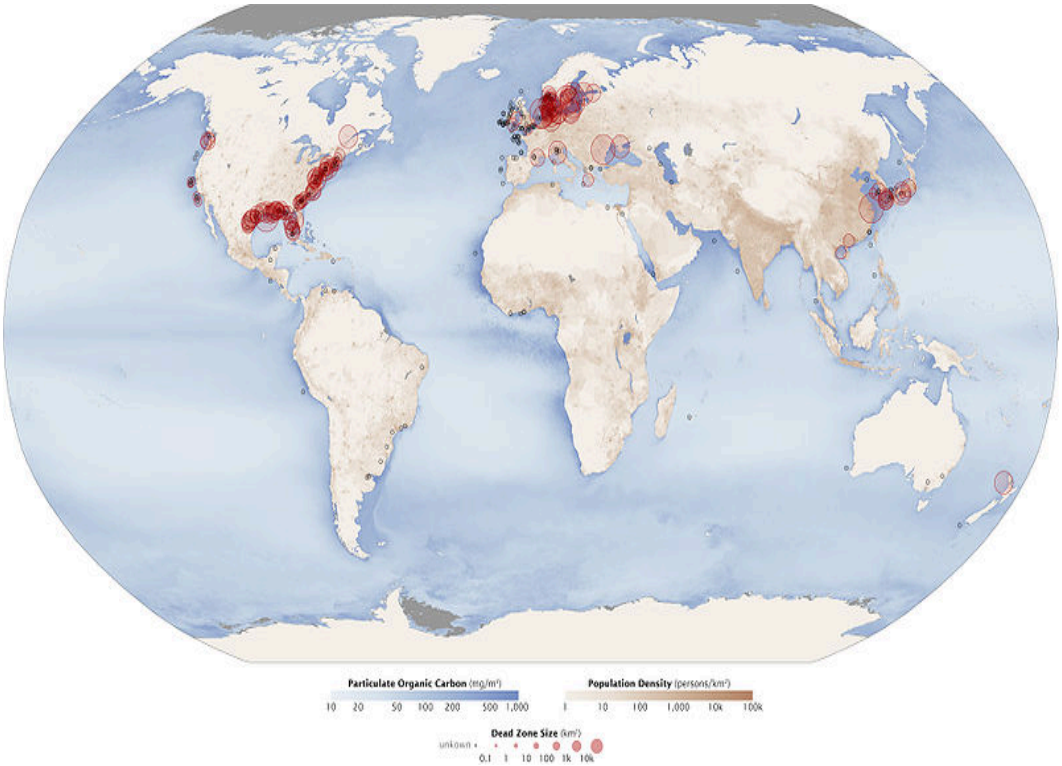


Figure 8.17 Aquatic Dead Zones Zones of hypoxia shown as red circles. Black dots show hypoxia zones of unknown size, brown shading shows population density, and blue shading shows density of particulate organic carbon, an indicator of organic productivity. *Source: Robert Simmon & Jesse Allen (http://commons.wikimedia.org/wiki/File:Aquatic_Dead_Zones.jpg) at NASA Earth Observatory via Wikimedia Commons*

Pathogens are disease-causing microorganisms, e.g., viruses, bacteria, parasitic worms, and protozoa, which cause a variety of intestinal diseases such as dysentery, typhoid fever, hepatitis, and cholera. Pathogens are the major cause of the water pollution crisis discussed at the beginning of this section. Unfortunately nearly a billion people around the world are exposed to waterborne pathogen pollution daily and around 1.5 million children mainly in underdeveloped countries die every year of waterborne diseases from pathogens. Pathogens enter water primarily from human and animal fecal waste

due to inadequate sewage treatment. In many underdeveloped countries, sewage is discharged into local waters either untreated or after only rudimentary treatment. In developed countries untreated sewage discharge can occur from overflows of combined sewer systems, poorly managed livestock factory farms, and leaky or broken sewage collection systems. Water with pathogens can be remediated by adding chlorine or ozone, by boiling, or by treating the sewage in the first place.

Oil spills are another kind of organic pollution. Oil spills can result from supertanker accidents such as the Exxon Valdez in 1989, which spilled 10 million gallons of oil into the rich ecosystem of offshore south Alaska and killed massive numbers of animals. The largest marine oil spill was the Deepwater Horizon disaster, which began with a natural gas explosion (see Figure **Figure 8.18**) at an oil well 65 km offshore of Louisiana and flowed for 3 months in 2010, releasing an estimated 200 million gallons of oil. The worst oil spill ever occurred during the Persian Gulf war of 1991, when Iraq deliberately dumped approximately 200 million gallons of oil in offshore Kuwait and set more than 700 oil well fires that released enormous clouds of smoke and acid rain for over nine months. During an oil spill on water, oil floats to the surface because it is less dense than water, and the lightest hydrocarbons evaporate, decreasing the size of the spill but polluting the air. Then, bacteria begin to decompose the remaining oil, in a process that can take many years. After several months only about 15% of the original volume may remain, but it is in thick asphalt lumps, a form that is particularly harmful to birds, fish, and shellfish. Cleanup operations can include *skimmer ships* that vacuum oil from the water surface (effective only for small spills), controlled *burning* (works only in early stages before the light, ignitable part evaporates but also pollutes the air), *dispersants* (detergents that break up oil to accelerate its decomposition, but some dispersants may be toxic to the ecosystem), and *bioremediation* (adding microorganisms that specialize in quickly decomposing oil, but this can disrupt the natural ecosystem).

Figure 8.18 Deepwater Horizon Explosion Boats fighting the fire from an explosion at the Deepwater Horizon drilling rig in Gulf of Mexico offshore Louisiana on April 20, 2010. *Source: United States Coast Guard (http://commons.wikimedia.org/wiki/File:Deepwater_Horizon_offshore_drilling_unit_on_fire_2010.jpg) via Wikimedia Commons*

Toxic chemicals involve many different kinds and sources, primarily from industry and mining. General kinds of toxic chemicals include *hazardous chemicals*, which are a wide variety of synthetic organic and inorganic chemicals such as acids, bases, cyanide, and a class of compounds called persistent organic pollutants that includes *DDT* (pesticide), *dioxin* (herbicide by-product), and *PCBs* (polychlorinated biphenyls, which were used as a liquid insulator in electric transformers). Persistent organic pollutants are long-lived in the environment, accumulate through the food chain (bioaccumulation), and can be toxic. Another category of toxic chemicals includes *radioactive materials* such as cesium, iodine, uranium, and radon gas, which can result in long-term exposure to radioactivity if it gets into the body. A final group of toxic chemicals is heavy metals such as lead, mercury, arsenic, cadmium, and chromium, which can accumulate

through the food chain. Heavy metals are commonly produced by industry and at metallic ore mines. Arsenic and mercury are discussed in more detail below. The US EPA regulates 83 contaminants in drinking water to ensure a safe public water supply. Similarly, at the international level the World Health Organization has drinking water standards for a variety of contaminants.

Arsenic (As) has been famous as an agent of death for many centuries. In large doses arsenic causes cancer and can be fatal. Only recently have scientists recognized that health problems can be caused by drinking small arsenic concentrations in water over a long time. It attacks the central nervous system and can damage the respiratory system, bladder, lungs, liver, and kidneys. It enters the water supply naturally from weathering of As-rich minerals and from human activities such as coal burning and smelting of metallic ores. The worst case of arsenic poisoning occurred in the densely populated impoverished country of Bangladesh, which had experienced 100,000s of deaths from diarrhea and cholera each year from drinking surface water contaminated with pathogens due to improper sewage treatment. In the 1970s the United Nations provided aid for millions of shallow water wells, which resulted in a dramatic drop in pathogenic diseases. Unfortunately, many of the wells produced water naturally rich in arsenic. Tragically, there are an estimated 77 million people (about half of the population) who inadvertently may have been exposed to toxic levels of arsenic in Bangladesh as a result. The World Health Organization has called it the largest mass poisoning of a population in history.

Mercury (Hg) is used in a variety of electrical products, such as dry cell batteries, fluorescent light bulbs, and switches, as well as in the manufacture of paint, paper, vinyl chloride, and fungicides. In the methylmercury form (CH_3Hg^+) it is highly toxic; ≥ 1 ppb of methylmercury represents water contaminated with mercury. Mercury and other toxic chemicals become concentrated in the food chain, especially in fish, in a process caused biological magnification (**Figure 8.19**). It acts on the central nervous system and can cause loss of sight, feeling, and hearing as well as nervousness, shakiness, and death. Like arsenic, mercury enters the water supply naturally from weathering of Hg-rich minerals and from human activities such as coal burning and metal processing. A famous mercury poisoning case in Minamata, Japan involved methylmercury-rich industrial discharge that caused high Hg levels in fish. People in the local fishing villages ate fish up to three times per day for over 30 years, which resulted in over 2,000 deaths. During that time the responsible company and national government did little to mitigate, help alleviate, or even acknowledge the problem.

Biological magnification represents the processes in an ecosystem that cause greater concentrations of a chemical, such as methylmercury, in organisms higher up the food chain. Mercury and methylmercury are present in only very small concentrations in seawater; however, at the base of the food chain algae absorb methylmercury. Then, small fish eat the algae, large fish and other organisms higher in the food chain eat the small fish, and so on. Fish and other aquatic organisms absorb methylmercury rapidly but eliminate it slowly from the body. Therefore, each step up the food chain increases the concentration from the step below (**Figure 8.19**). Largemouth bass can concentrate methylmercury up to 10 million times over the water concentration and fish-eating birds can concentrate it even higher. Other chemicals that exhibit biological magnification are DDT, PCBs, and arsenic.

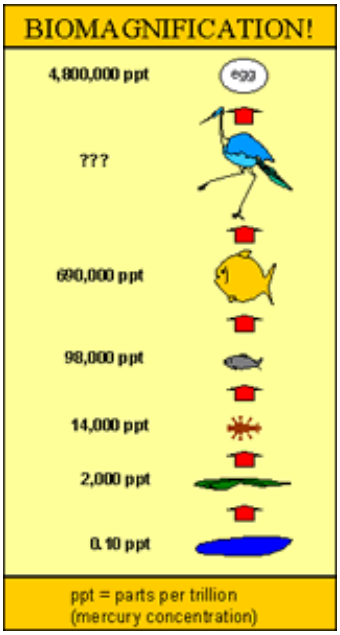


Figure 8.19 Biological magnification An illustrative example of biological magnification of mercury from water through the food chain and into a bird's egg. *Source:* ***U.S. Geological Survey (http://sofia.usgs.gov/sfrsf/rooms/mercury/achilles_heel/cause.html)***

Other water pollutants include sediment and heat. Muddy water is bad for drinking but even worse for underwater plants that need sunlight for photosynthesis. Much of the sediment in water bodies is derived from the erosion of soil, so it also represents a loss of agricultural productivity. Thermal pollution involves the release of heated waters from power plants and industry to surface water, causing a drop in the dissolved O_2 content, which can stress fish.

Hard water contains abundant calcium and magnesium, which reduces its ability to develop soapsuds and enhances *scale* (calcium and magnesium carbonate minerals) formation on hot water equipment. Water softeners remove calcium and magnesium, which allows the water to lather easily and resist scale formation. Hard water develops naturally from the dissolution of calcium and magnesium carbonate minerals in soil; it does not have negative health effects in people.

Groundwater pollution can occur from underground sources and all of the pollution sources that contaminate surface waters. Common sources of groundwater pollution are leaking underground storage tanks for fuel, septic tanks, agricultural activity, and landfills. Common groundwater pollutants include nitrate, pesticides, volatile organic compounds, and petroleum products. Polluted groundwater can be a more serious problem than polluted surface water because the pollution in groundwater may go undetected for a long time because usually it moves very slowly. As a result, the pollution in groundwater may create a contaminant plume, a large body of flowing polluted groundwater, making cleanup very costly. By the time groundwater contamination is detected, the entity responsible for the pollution may be bankrupt or nonexistent. Another troublesome feature of groundwater pollution is that small amounts of certain pollutants, e.g., petroleum products and organic solvents, can contaminate large areas. In Denver, Colorado 80 liters of several organic solvents contaminated 4.5 trillion liters of groundwater and produced a 5 km long contaminant plume. Most groundwater contamination occurs in shallow, unconfined aquifers located near the contamination source. Confined aquifers are less susceptible to pollution from the surface because of protection by the confining layer. A major threat to groundwater quality is from underground fuel storage tanks. Fuel tanks commonly are stored underground at gas stations to reduce explosion hazards. Before 1988 in the U.S. these storage tanks could be made of metal, which can corrode, leak, and quickly contaminate local groundwater. Now, leak detectors are required and the metal storage tanks are supposed to be protected from corrosion or replaced with fiberglass tanks. Currently there are around 600,000 underground fuel storage tanks in the U.S. and over 30% still do not comply with EPA regulations regarding either release prevention or leak detection.

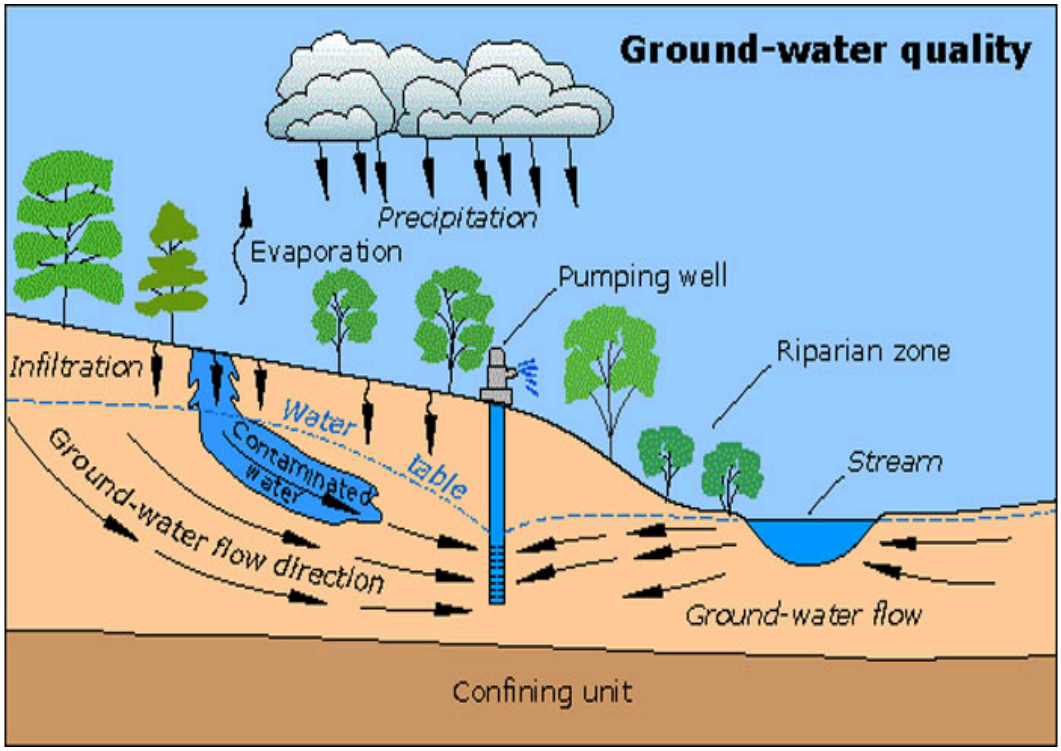


Figure 8.20 Contaminant Plume in Groundwater Mapping how a contaminant plume will migrate once it reaches groundwater requires understanding of the pollutant's chemical properties, local soil characteristics, and how permeable the aquifer is. *Source: **United States Geological Survey (http://www.learner.org/courses/envsci/visual/visual.php?shortname=contaminant_flow)***

Sustainable Solutions to the Water Pollution Crisis?

Resolution of the global water pollution crisis described at the beginning of this section requires multiple approaches to improve the quality of our fresh water and move towards sustainability. The most deadly form of water pollution, pathogenic microorganisms that cause waterborne diseases, kills almost 2 million people in underdeveloped countries every year. The best strategy for addressing this problem is proper sewage (wastewater) treatment. Untreated sewage is not only a major cause of pathogenic diseases, but also a major source of other pollutants, including oxygen-demanding waste, plant nutrients (N and P), and toxic heavy metals. Wastewater treatment is done at a sewage treatment plant in urban areas and through a septic tank system in rural areas.

The main purpose of a sewage treatment plant is to remove organic matter (oxygen-demanding waste) and kill bacteria; special methods also can be used to remove plant nutrients and other pollutants. The numerous processing steps at a conventional sewage treatment plant (**Figure 8.21**) include *pretreatment* (screening and removal of sand and gravel), *primary treatment* (settling or floatation to remove organic solids, fat, and grease), *secondary treatment* (aerobic bacterial decomposition of organic solids), *tertiary treatment* (bacterial decomposition of nutrients and filtration), *disinfection* (treatment with chlorine, ozone, ultraviolet light, or bleach), and either *discharge* to surface waters (usually a local river) or *reuse* for some other purpose, such as irrigation, habitat preservation, and artificial groundwater recharge. The concentrated organic solid produced during primary and secondarytreatment is called sludge, which is treated in a variety of ways including landfill disposal, incineration, use as fertilizer, and anaerobic bacterial decomposition, which is done in the absence of oxygen. Anaerobic decomposition of sludge produces methane gas, which can be used as an energy source. To reduce water pollution problems, separate sewer systems (where street runoff goes to rivers and only wastewater goes to a wastewater treatment plant) are much better than combined sewer systems, which can overflow and release untreated sewage into surface waters during heavy rain. Some cities such as Chicago, Illinois have constructed large underground caverns and also use abandoned rock quarries to hold storm sewer overflow. After the rain stops, the stored water goes to the sewage treatment plant for processing.

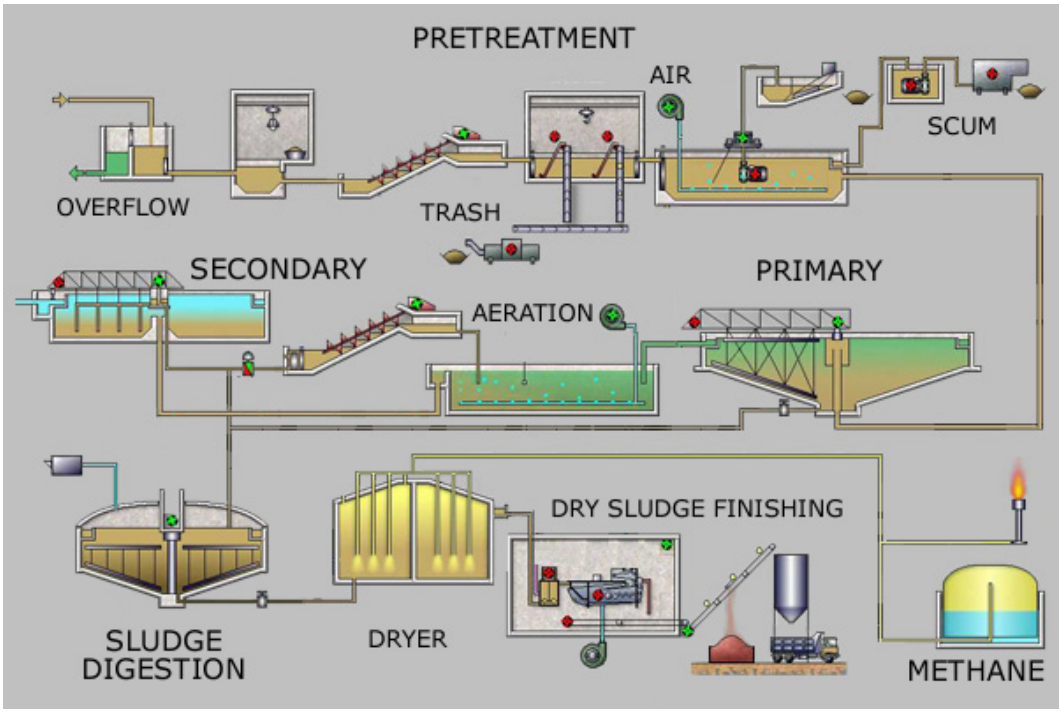


Figure 8.21 Steps at a Sewage Treatment Plant The numerous processing steps at a conventional sewage treatment plant include pretreatment (screening and removal of sand and gravel), primary treatment (settling or floatation to remove organic solids, fat, and grease), secondary treatment (aerobic bacterial decomposition of organic solids), tertiary treatment (bacterial decomposition of nutrients and filtration), disinfection (treatment with chlorine, ozone, ultraviolet light, or bleach), and either discharge to surface waters (usually a local river) or reuse for some other purpose, such as irrigation, habitat preservation, and artificial groundwater recharge. *Source: **Leonard G. (http://en.wikipedia.org/wiki/File:ESQUEMPEQUE-EN.jpg)** via Wikipedia*

A septic tank system is an individual sewage treatment system for homes in rural and even some urban settings. The basic components of a septic tank system (**Figure 8.22** include a sewer line from the house, a *septic tank* (a large container where sludge settles to the bottom and microorganisms decompose the organic solids anaerobically), and the drain field (network of perforated pipes where the clarified water seeps into the soil and is further purified by bacteria). Water pollution problems occur if the septic tank malfunctions, which usually occurs when a system is established in the wrong type of soil or maintained poorly.

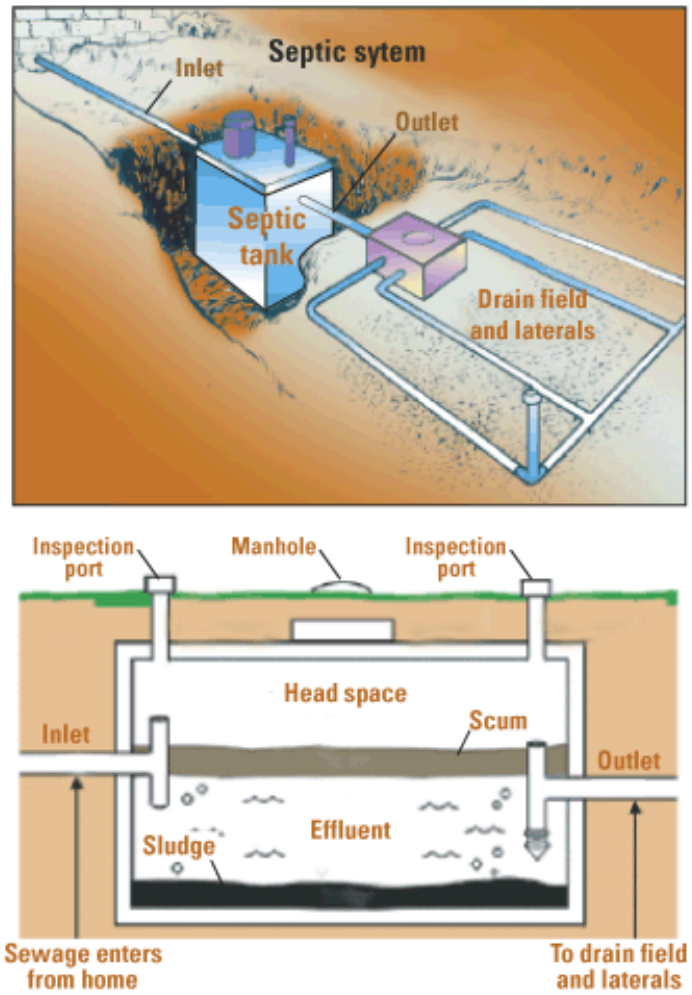


Figure 8.22 Septic System Septic tank system for sewage treatment. *Source: United States Geological Survey (http://pubs.usgs.gov/fs/fs07203/)*

For many developing countries, financial aid is necessary to build adequate sewage treatment facilities; however, the World Health Organization estimates an estimated cost savings of between $3 and $34 for every $1 invested in clean water delivery and sanitation (**Water for Life, 2005**). The cost savings are from health care savings, gains in work and school productivity, and deaths prevented. Simple and inexpensive techniques for treating water at home include chlorination, filters, and solar disinfection. Another alternative is to use constructed wetlands technology (marshes built to treat contaminated water), which is simpler and cheaper than a conventional sewage treatment plant.

Bottled water is *not* a sustainable solution to the water crisis, despite exponential growth in popularity in the U.S. and the world. Bottled water is not necessarily any safer than the U.S. public water supply, it costs on average about 700 times more than U.S. tap water, and every year it uses approximately 200 billion plastic and glass bottles that have a relatively low rate of recycling. Compared to tap water, it uses much more energy, mainly in bottle manufacturing and long-distance transportation. If you don't like the taste of your tap water, then please use a water filter instead of bottled water!

Additional sustainable solutions to the water pollution crisis include legislation to eliminate or greatly reduce point sources of water pollution. In the U.S., the Clean Water Act of 1972 and later amendments led to major improvements in water quality (**m47268 (http://legacy.cnx.org/content/m47268/1.5/#id1172489677600)**). Nonpoint sources of water pollution, e.g., agricultural runoff and urban runoff, are much harder to regulate because of their widespread, diffuse nature. There are many construction and agricultural practices that reduce polluted runoff including no-till farming and sediment traps. Artificial aeration or mechanical mixing can remediate lakes with oxygen depletion. Specific things that we can do to reduce urban runoff include the following: keep soil, leaves, and grass clippings off driveways, sidewalks, and streets; don't pour used motor oil, antifreeze, paints, pesticides, or any household hazardous chemical down the storm sewer or drain; recycle used motor oil; use hazardous waste disposal programs offered by the community; compost your organic waste; don't use fertilizers and herbicides on your lawn; and flush pet waste down the toilet.

During the early 1900s rapid industrialization in the U.S. resulted in widespread water pollution due to free discharge of waste into surface waters. The Cuyahoga River in northeast Ohio caught fire numerous times (**Figure 8.23**), including a famous fire in 1969 that caught the nation's attention. In 1972 Congress passed one of the most important environmental laws in U.S. history, the Federal Water Pollution Control Act, which is more commonly called the Clean Water Act. The purpose of the Clean Water Act and later amendments is to maintain and restore water quality, or in simpler terms to make our water swimmable and fishable. It became illegal to dump pollution into surface water unless there was formal permission and U.S. water quality improved significantly as a result. More progress is needed because currently the EPA considers over 40,000 U.S. water bodies as impaired, most commonly due to pathogens, metals, plant nutrients, and oxygen depletion. Another concern is protecting groundwater quality, which is not yet addressed sufficiently by federal law.

Figure 8.23 Cuyahoga River on Fire*Source:* ***National Oceanic and Atmospheric (http://oceanservice.noaa.gov/education/kits/pollution/media/supp_pol02d.html)***

Sometimes slow flow through a soil can naturally purify groundwater because some pollutants, such as P, pesticides, and heavy metals, chemically bind with surfaces of soil clays and iron oxides. Other pollutants are not retained by soil particles: These include N, road salt, gasoline fuel, the herbicide atrazine, tetrachloroethylene (a carcinogenic cleaning solvent used in dry cleaning), and vinyl chloride. In other cases, slow groundwater flow can allow bacteria to decompose dead organic matter and certain pesticides. There are many other ways to remediate polluted groundwater. Sometimes the best solution is to stop the pollution source and allow natural cleanup. Specific treatment methods depend on the geology, hydrology, and pollutant because some light contaminants flow on top of groundwater, others dissolve and flow with groundwater, and dense contaminants can sink below groundwater. A common cleanup method called pump and treat involves pumping out the contaminated groundwater and treating it by oxidation, filtration, or biological methods. Sometimes soil must be excavated and sent to a landfill. In-situ treatment methods include adding chemicals to immobilize heavy metals, creating a permeable reaction zone with metallic iron that can destroy organic solvents, or using bioremediation by adding oxygen or nutrients to stimulate growth of microorganisms.

References

Water for Life: Making it Happen (2005) World Health Organization and UNICEF. Retrieved from **http://www.who.int/water_sanitation_health/waterforlife.pdf (http://www.who.int/water_sanitation_health/waterforlife.pdf)**

World Health Statistics (2010) World Health Organization. Retrieved from **http://www.who.int/whosis/whostat/EN_WHS10_Full.pdf (http://www.who.int/whosis/whostat/EN_WHS10_Full.pd)**

9 | INTRODUCTION TO CELL BIOLOGY

9.1 | Introduction to Cells

Introduction

> “The history of the knowledge of the phenomena of life and of the organized world can be divided into two main periods. For a long time anatomy, and particularly the anatomy of the human body, was the alpha and omega of scientific knowledge. Further progress only became possible with the discovery of the microscope. A long time had yet to pass until through Schwann the **cell** was established as the final biological unit. It would mean bringing coals to Newcastle were I to describe here the immeasurable progress which biology, in all its branches, owes to the introduction of this concept of the cell. For this concept is the axis around which the whole of modern science revolves.”

Paul Ehrlich, "Partial Cell Functions", Nobel Lecture, December 11, 1908

Ehrlich's enthusiasm for the cell is understandable. A single cell is the basic unit of life, and the starting point for each and every human and other organism on the planet. A cell is the smallest unit of a living thing. A living thing, whether made of one cell (like bacteria) or many cells (like a human), is called an organism. Thus, cells are the basic building blocks of all organisms, and the study of cells is at the very heart of the research enterprise that we call biological science.

Several cells of one kind that interconnect with each other and perform a shared function form tissues, several tissues combine to form an organ (your stomach, heart, or brain), and several organs make up an organ system (such as the digestive system, circulatory system, or nervous system). Several systems that function together form an organism (like a human being). Here, we will examine the structure and function of cells.

There are many types of cells, all grouped into one of two broad categories: prokaryotic and eukaryotic. For example, both animal and plant cells are classified as eukaryotic cells, whereas bacterial cells are classified as prokaryotic. Before discussing the criteria for determining whether a cell is prokaryotic or eukaryotic, let’s first examine how biologists study cells.

Microscopy

Cells vary in size. With few exceptions, individual cells cannot be seen with the naked eye, so scientists use microscopes (micro- = “small”; -scope = “to look at”) to study them. A **microscope** is an instrument that magnifies an object. Most photographs of cells are taken with a microscope, and these images can also be called micrographs.

The optics of a microscope’s lenses change the orientation of the image that the user sees. A specimen that is right-side up and facing right on the microscope slide will appear upside-down and facing left when viewed through a microscope, and vice versa. Similarly, if the slide is moved left while looking through the microscope, it will appear to move right, and if moved down, it will seem to move up. This occurs because microscopes use two sets of lenses to magnify the image. Because of the manner by which light travels through the lenses, this system of two lenses produces an inverted image

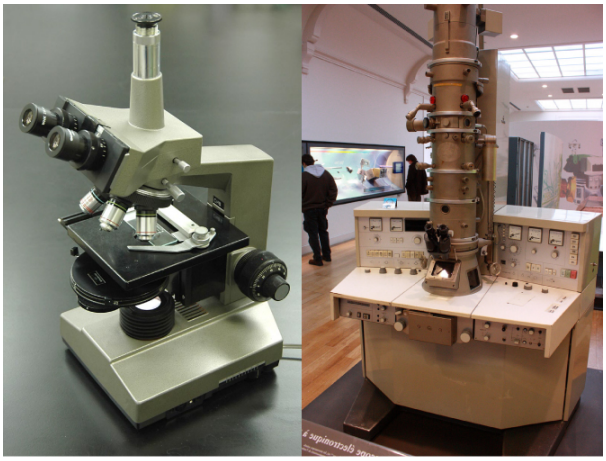
(binocular, or dissecting microscopes, work in a similar manner, but include an additional magnification system that makes the final image appear to be upright).

Light Microscopes

To give you a sense of cell size, a typical human red blood cell is about eight millionths of a meter or eight micrometers (abbreviated as eight μm, or eight μ) in diameter; the head of a pin of is about two thousandths of a meter (two mm) in diameter. That means about 250 red blood cells could fit on the head of a pin.

Most student microscopes are classified as **light microscopes** (**Figure 9.1a**). Visible light passes and is bent through the lens system to enable the user to see the specimen. Light microscopes are advantageous for viewing living organisms, but since individual cells are generally transparent, their components are not distinguishable unless they are colored with special stains. Staining, however, usually kills the cells.

Light microscopes commonly used in the undergraduate college laboratory magnify up to approximately 400 times. Two parameters that are important in microscopy are magnification and resolving power. Magnification is the process of enlarging an object in appearance. Resolving power is the ability of a microscope to distinguish two adjacent structures as separate: the higher the resolution, the better the clarity and detail of the image. When oil immersion lenses are used for the study of small objects, magnification is usually increased to 1,000 times. In order to gain a better understanding of cellular structure and function, scientists typically use electron microscopes.



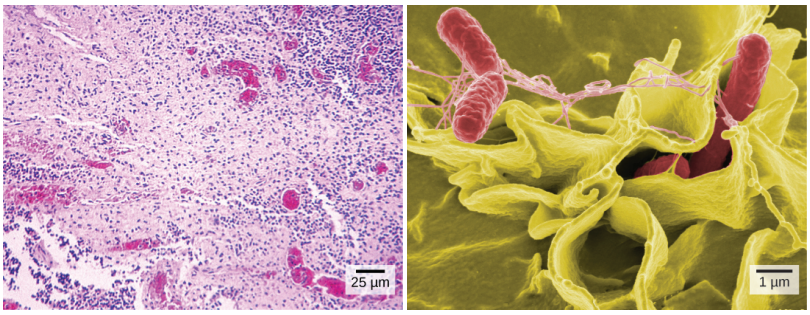
(a) (b)

Figure 9.1 (a) Most light microscopes used in a college biology lab can magnify cells up to approximately 400 times and have a resolution of about 200 nanometers. (b) Electron microscopes provide a much higher magnification, 100,000x, and a have a resolution of 50 picometers. (credit a: modification of work by "GcG"/Wikimedia Commons; credit b: modification of work by Evan Bench)

Electron Microscopes

In contrast to light microscopes, **electron microscopes** (**Figure 9.1b**) use a beam of electrons instead of a beam of light. Not only does this allow for higher magnification and, thus, more detail (**Figure 9.2**), it also provides higher resolving power. The method used to prepare the specimen for viewing with an electron microscope kills the specimen. Electrons have short wavelengths (shorter than photons) that move best in a vacuum, so living cells cannot be viewed with an electron microscope.

In a scanning electron microscope, a beam of electrons moves back and forth across a cell's surface, creating details of cell surface characteristics. In a transmission electron microscope, the electron beam penetrates the cell and provides details of a cell's internal structures. As you might imagine, electron microscopes are significantly more bulky and expensive than light microscopes.



(a) (b)

Figure 9.2 (a) These *Salmonella* bacteria appear as tiny purple dots when viewed with a light microscope. (b) This scanning electron microscope micrograph shows *Salmonella* bacteria (in red) invading human cells (yellow). Even though subfigure (b) shows a different *Salmonella* specimen than subfigure (a), you can still observe the comparative increase in magnification and detail. (credit a: modification of work by CDC/Armed Forces Institute of Pathology, Charles N. Farmer, Rocky Mountain Laboratories; credit b: modification of work by NIAID, NIH; scale-bar data from Matt Russell)

Cell Theory

The microscopes we use today are far more complex than those used in the 1600s by Antony van Leeuwenhoek, a Dutch shopkeeper who had great skill in crafting lenses. Despite the limitations of his now-ancient lenses, van Leeuwenhoek observed the movements of protista (a type of single-celled organism) and sperm, which he collectively termed “animalcules.”

In a 1665 publication called *Micrographia*, experimental scientist Robert Hooke coined the term “cell” for the box-like structures he observed when viewing cork tissue through a lens. In the 1670s, van Leeuwenhoek discovered bacteria and protozoa. Later advances in lenses, microscope construction, and staining techniques enabled other scientists to see some components inside cells.

By the late 1830s, botanist Matthias Schleiden and zoologist Theodor Schwann were studying tissues and proposed the **unified cell theory**, which states that all living things are composed of one or more cells, the cell is the basic unit of life, and new cells arise from existing cells. Rudolf Virchow later made important contributions to this theory.

career CONNECTION

Cytotechnologist

Have you ever heard of a medical test called a Pap smear (Figure 9.3)? In this test, a doctor takes a small sample of cells from the uterine cervix of a patient and sends it to a medical lab where a cytotechnologist stains the cells and examines them for any changes that could indicate cervical cancer or a microbial infection.

Cytotechnologists (cyto- = "cell") are professionals who study cells via microscopic examinations and other laboratory tests. They are trained to determine which cellular changes are within normal limits and which are abnormal. Their focus is not limited to cervical cells; they study cellular specimens that come from all organs. When they notice abnormalities, they consult a pathologist, who is a medical doctor who can make a clinical diagnosis.

Cytotechnologists play a vital role in saving people's lives. When abnormalities are discovered early, a patient's treatment can begin sooner, which usually increases the chances of a successful outcome.

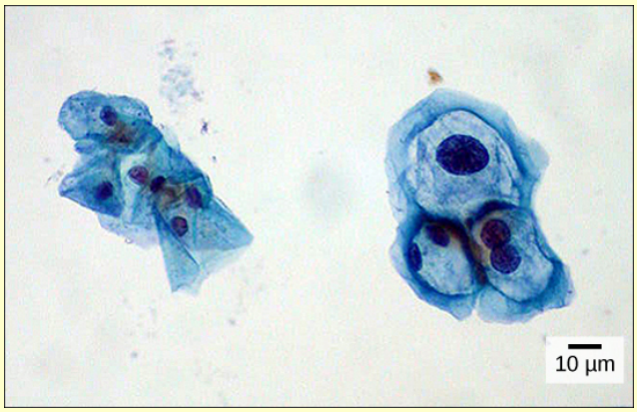


Figure 9.3 These uterine cervix cells, viewed through a light microscope, were obtained from a Pap smear. Normal cells are on the left. The cells on the right are infected with human papillomavirus (HPV). Notice that the infected cells are larger; also, two of these cells each have two nuclei instead of one, the normal number. (credit: modification of work by Ed Uthman, MD; scale-bar data from Matt Russell)

The Phylogenetic Relationship of Life

All of life can be grouped into three Domains Archaea, Bacteria and Eukarya (**Figure 9.4**). Even though Archaea and Bacteria are prokaryotes, there are enough differences between Archaea and Bacteria that warrant them being in different Domains. Within the Domain Eukarya, there are at least four Kingdoms Protists (multiple kingdoms), Fungi, Plantae and Animalia. Figure 4 shows the current phylogentic tree of all these various groups that we will be exploring.

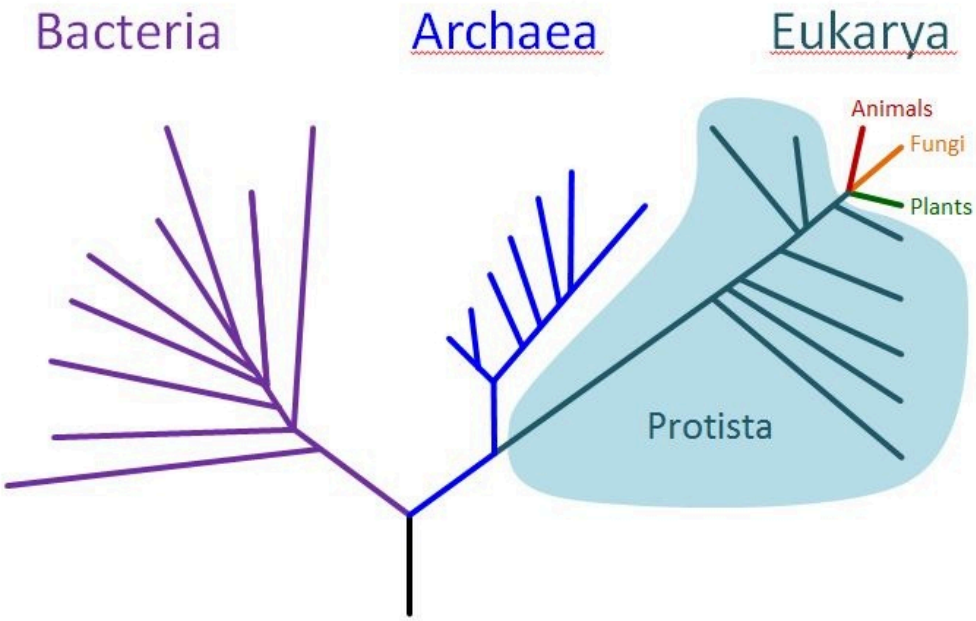


Figure 9.4 The phylogenetic tree of the Domains Bacteria, Archeae and the four Kingdoms of Eukarya. Work by Eva Horne

Our approach to Cell Biology

In this textbook, we explore cells and the chemistry of life in the reverse order of many traditional textbooks. First, we explore the diversity of life at the cellular level and then we investigate the chemistry of life or Biochemistry. The reason why we take this approach is based on the fact that many students have had some exposure to cells and this exposure allows them to connect with the material. In addition, we feel a macro to micro approach to investigating this material allows the learner the ability to better connect how the biochemistry relates to the functioning cell.

9.2 | Prokaryotic Cells

Introduction

> "With the identification and characterization of the kingdoms we are for the first time beginning to see the overall phylogenetic structure of the living world. It is not structured in a bipartite way along the lines of the organizationally dissimilar prokaryote and eukaryote. Rather, it is (at least) tripartite, comprising (i) the typical bacteria, (ii) the line of descent manifested in eukaryotic cytoplasms, and (iii) a little explored grouping, represented so far only by methanogenic bacteria."

Cal Woese and George Fox, "Phylogenetic structure of the prokaryotic domain: the primary kingdoms", *Proceedings of the National Academy of Science, USA*. 1977 74(11):5088-90.

Cells fall into one of two broad categories: prokaryotic and eukaryotic. Until the publication of the paper cited above, it was believed that the prokaryotic cells were all members of the group we know as Bacteria. Woese and Fox (and others) laid out a strong case for another class of prokaryotic cells, which they christened the Archaea. Cells of animals, plants, fungi, and

protists are all eukaryotes (eu- = “true”, karyo = "nucleus") and are made up of eukaryotic cells. The single-celled organisms of the domains Bacteria and Archaea are prokaryotes (pro- = “before”). We will consider these two domains here, since they have some similarities, but it is good to remember that Bacteria and Archea are as different from each other as they are different from the Eukaryotes.

Components of Prokaryotic Cells

All cells share four common components: 1) a plasma membrane, an outer covering that separates the cell’s interior from its surrounding environment; 2) cytoplasm, consisting of a jelly-like cytosol within the cell in which other cellular components are found; 3) DNA, the genetic material of the cell; and 4) ribosomes, which synthesize proteins. However, prokaryotes differ from eukaryotic cells in several ways.

A **prokaryote** is a simple, mostly single-celled (unicellular) organism that lacks a nucleus, or any other membrane-bound organelle. We will shortly come to see that this is significantly different in eukaryotes. Prokaryotic DNA is found in a central part of the cell: the **nucleoid** (Figure 9.5).

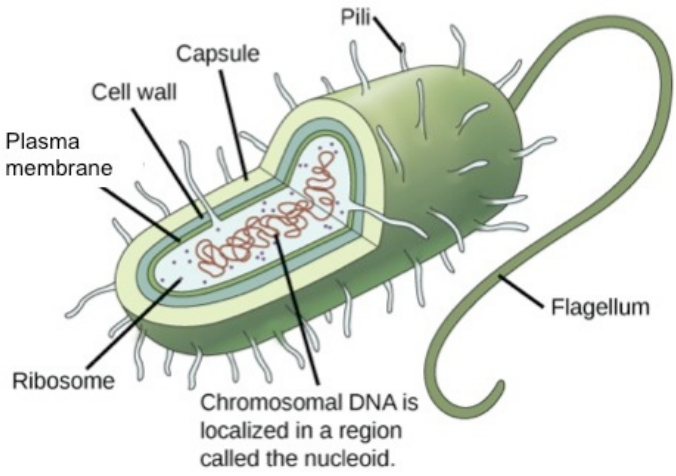


Figure 9.5 This figure shows the generalized structure of a prokaryotic cell. All prokaryotes have chromosomal DNA localized in a nucleoid, ribosomes, a plasma membrane, and a cell wall. The other structures shown are present in some, but not all, bacteria.

Most prokaryotes have a peptidoglycan cell wall and many have a polysaccharide capsule (Figure 9.5). The cell wall acts as an extra layer of protection, helps the cell maintain its shape, and prevents dehydration. The capsule enables the cell to attach to surfaces in its environment. Some prokaryotes have flagella, pili, or fimbriae. Flagella are used for locomotion. Pili are used to exchange genetic material during a type of reproduction called conjugation. Fimbriae are used by bacteria to attach to a host cell.

career CONNECTION

Microbiologist

The most effective action anyone can take to prevent the spread of contagious illnesses is to wash his or her hands. Why? Because microbes (organisms so tiny that they can only be seen with microscopes) are ubiquitous. They live on doorknobs, money, your hands, and many other surfaces. If someone sneezes into his hand and touches a doorknob, and afterwards you touch that same doorknob, the microbes from the sneezer's mucus are now on your hands. If you touch your hands to your mouth, nose, or eyes, those microbes can enter your body and could make you sick.

However, not all microbes (also called microorganisms) cause disease; most are actually beneficial. You have microbes in your gut that make vitamin K. Other microorganisms are used to ferment beer and wine.

Microbiologists are scientists who study microbes. Microbiologists can pursue a number of careers. Not only do they work in the food industry, they are also employed in the veterinary and medical fields. They can work in the pharmaceutical sector, serving key roles in research and development by identifying new sources of antibiotics that could be used to treat bacterial infections.

Environmental microbiologists may look for new ways to use specially selected or genetically engineered microbes for the removal of pollutants from soil or groundwater, as well as hazardous elements from contaminated sites. These uses of microbes are called bioremediation technologies. Microbiologists can also work in the field of bioinformatics, providing specialized knowledge and insight for the design, development, and specificity of computer models of, for example, bacterial epidemics.

Cell Size

At 0.1 to 5.0 μm in diameter, prokaryotic cells are significantly smaller than eukaryotic cells, which have diameters ranging from 10 to 100 μm (**Figure 9.6**). The small size of prokaryotes allows ions and organic molecules that enter them to quickly diffuse to other parts of the cell. Similarly, any wastes produced within a prokaryotic cell can quickly diffuse out. This is not the case in eukaryotic cells, which have developed different structural adaptations to enhance intracellular transport.

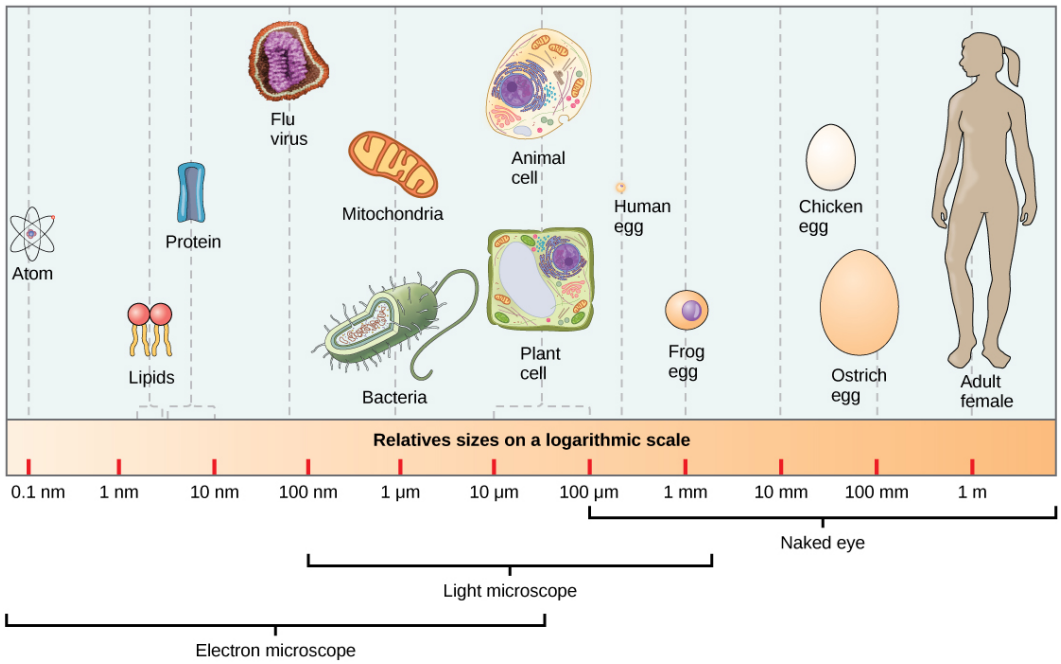


Figure 9.6 This figure shows relative sizes of microbes on a logarithmic scale (recall that each unit of increase in a logarithmic scale represents a 10-fold increase in the quantity being measured).

Small size, in general, is necessary for all cells, whether prokaryotic or eukaryotic. Let's examine why that is so. First, we'll consider the area and volume of a typical cell. Not all cells are spherical in shape, but most tend to approximate a sphere.

You may remember from your high school geometry course that the formula for the surface area of a sphere is $4\pi r^2$, while the formula for its volume is $4\pi r^3/3$. Thus, as the radius of a cell increases, its surface area increases as the square of its radius, but its volume increases as the cube of its radius (much more rapidly). Therefore, as a cell increases in size, its surface area-to-volume ratio decreases. This same principle would apply if the cell had the shape of a cube (**Figure 9.7**). If the cell grows too large, the plasma membrane will not have sufficient surface area to support the rate of diffusion required for the increased volume. In other words, as a cell grows, it becomes less efficient. One way to become more efficient is to divide; another way is to develop organelles that perform specific tasks. These adaptations lead to the development of more sophisticated cells called eukaryotic cells.

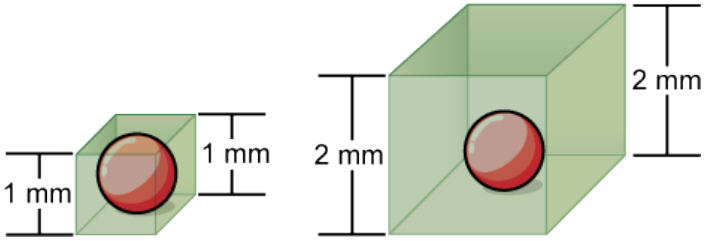


Figure 9.7 Notice that as a cell increases in size, its surface area-to-volume ratio decreases. When there is insufficient surface area to support a cell's increasing volume, a cell will either divide or die. The cell on the left has a volume of 1 mm^3 and a surface area of 6 mm^2, with a surface area-to-volume ratio of 6 to 1, whereas the cell on the right has a volume of 8 mm^3 and a surface area of 24 mm^2, with a surface area-to-volume ratio of 3 to 1.

9.3 | Eukaryotic Cells

Introduction

> "The cell, too, has a geography. And its reactions occur in a colloidal apparatus, of which the form, and the catalytic activity of its manifold surfaces, must efficiently contribute to the due guidance of chemical reactions. "

Frederick C. Hopkins, "Some Aspects of Biochemistry", *The Irish Journal of Medical Science*, 1932, 79:344

Have you ever heard the phrase "form follows function?" It's a philosophy practiced in many industries. In architecture, this means that buildings should be constructed to support the activities that will be carried out inside them. For example, a skyscraper should be built with several elevator banks; a hospital should be built so that its emergency room is easily accessible.

Our natural world also utilizes the principle of form following function, especially in cell biology, and this will become clear as we explore eukaryotic cells (**Figure 9.8**). Unlike prokaryotic cells, **eukaryotic cells** have: 1) a membrane-bound nucleus; 2) numerous membrane-bound **organelles** such as the endoplasmic reticulum, Golgi apparatus, chloroplasts, mitochondria, and others; and 3) several, rod-shaped chromosomes. Of these, Hopkins points out the membrane systems (the "manifold surfaces") as being especially important. Indeed, the study of eukaryotic cells is in many respects the study of the structure and function of these surfaces. Different organelles have different forms and functions based on the form and function of their component membranes. Indeed, the word "organelle" means "little organ," and organelles have specialized shapes and specialized functions, just as the organs of your body have specialized shapes and functions.

At this point, it should be clear to you that eukaryotic cells have a more complex structure than prokaryotic cells. Organelles allow different functions to be compartmentalized in different areas of the cell. Before turning to organelles, let's first examine two important components of the cell: the plasma membrane and the cytoplasm.

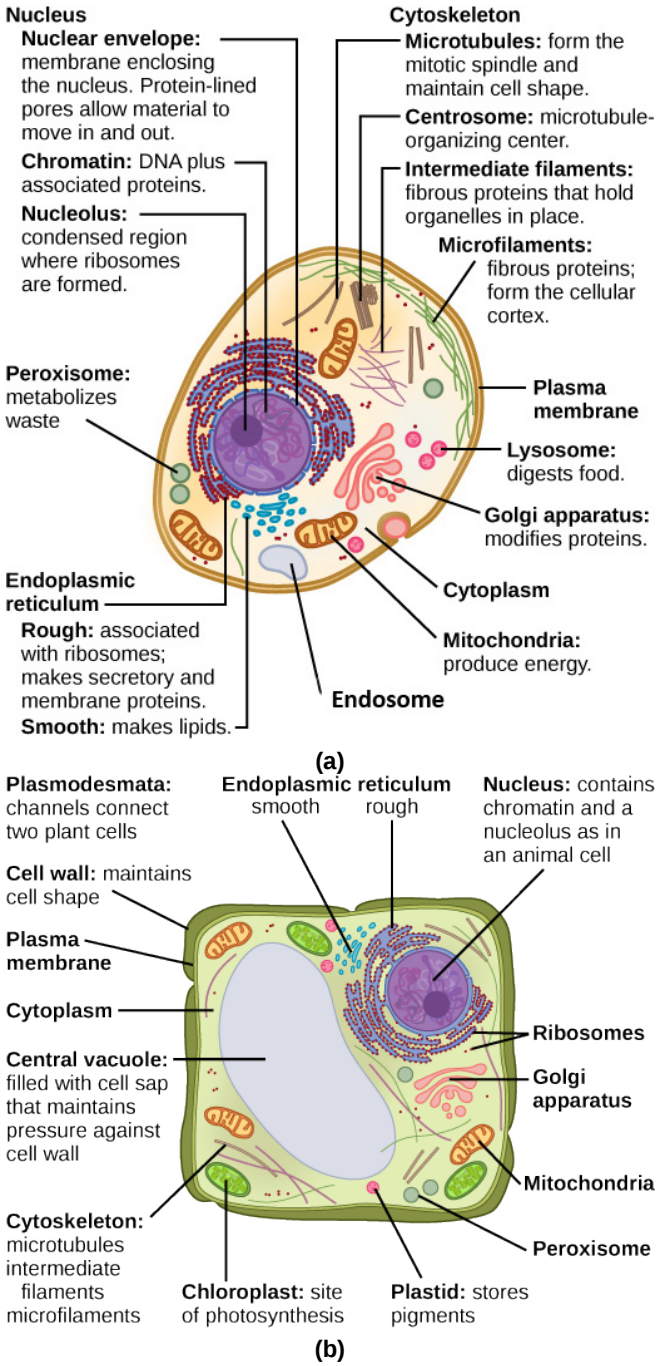


Figure 9.8 These figures show the major organelles and other cell components of (a) a typical animal cell and (b) a typical eukaryotic plant cell. The plant cell has a cell wall, chloroplasts, plastids, and a central vacuole—structures not found in animal cells. Plant cells do not have lysosomes or centrosomes.

The Plasma Membrane

Like prokaryotes, eukaryotic cells have a **plasma membrane** (**Figure 9.9**), a phospholipid bilayer with embedded proteins that separates the internal contents of the cell from its surrounding environment. A phospholipid is a lipid molecule with two fatty acid chains and a phosphate-containing group. The plasma membrane controls the passage of organic molecules, ions, water, and oxygen into and out of the cell. Wastes (such as carbon dioxide and ammonia) also leave the cell by passing through the plasma membrane. Thus, the plasma membrane is said to be **semi-permeable**

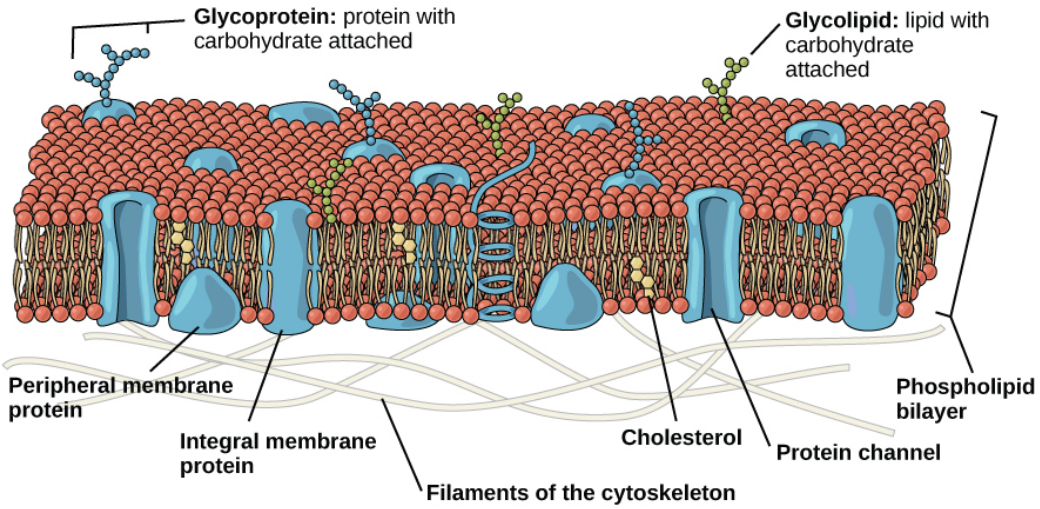


Figure 9.9 The eukaryotic plasma membrane is a phospholipid bilayer with proteins and cholesterol embedded in it.

The plasma membranes of cells that specialize in absorption are folded into fingerlike projections called microvilli (singular = microvillus); (**Figure 9.10**). Such cells are typically found lining the small intestine, the organ that absorbs nutrients from digested food. This is an excellent example of form following function. People with celiac disease have an immune response to gluten, which is a protein found in wheat, barley, and rye. The immune response damages microvilli, and thus, afflicted individuals cannot absorb nutrients. This leads to malnutrition, cramping, and diarrhea. Patients suffering from celiac disease must follow a gluten-free diet.

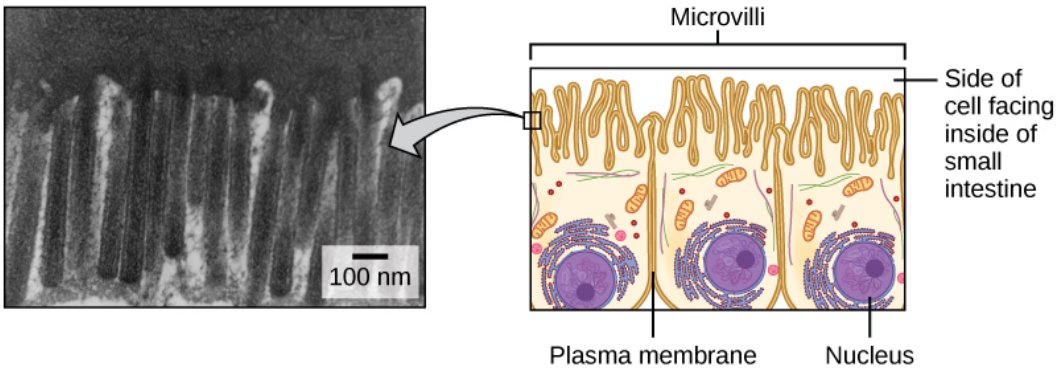


Figure 9.10 Microvilli, shown here as they appear on cells lining the small intestine, increase the surface area available for absorption. These microvilli are only found on the area of the plasma membrane that faces the cavity from which substances will be absorbed. (credit "micrograph": modification of work by Louisa Howard)

The Cytoplasm

The **cytoplasm** is the entire region of a cell between the plasma membrane and the nuclear envelope (a structure to be discussed shortly). It is made up of organelles suspended in the gel-like **cytosol**, the cytoskeleton, and various chemicals (**Figure 9.8**). Even though the cytoplasm consists of 70 to 80 percent water, it has a semi-solid consistency, which comes from the proteins within it. However, proteins are not the only organic molecules found in the cytoplasm. Glucose and other simple sugars, polysaccharides, amino acids, nucleic acids, fatty acids, and derivatives of glycerol are found there, too. Ions of sodium, potassium, calcium, and many other elements are also dissolved in the cytoplasm. Many metabolic reactions, including protein synthesis, take place in the cytoplasm.

The Nucleus

Typically, the nucleus is the most prominent organelle in a cell (**Figure 9.8**). The **nucleus** (plural = nuclei) houses the cell's DNA and directs the synthesis of ribosomes and proteins. Let's look at it in more detail (**Figure 9.11**).

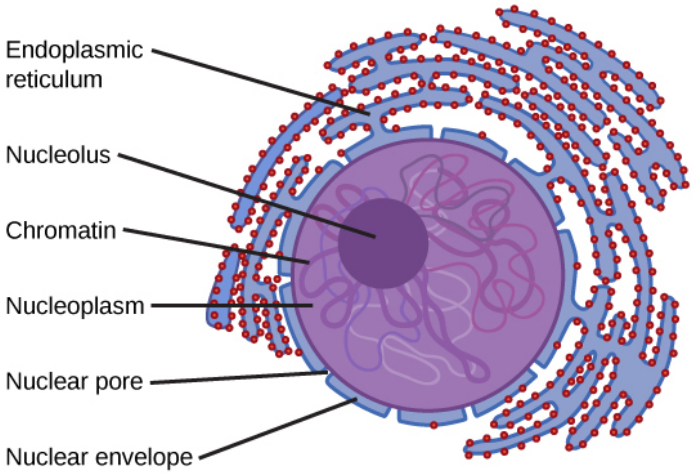


Figure 9.11 The nucleus stores chromatin (DNA plus proteins) in a gel-like substance called the nucleoplasm. The nucleolus is a condensed region of chromatin where ribosome synthesis occurs. The boundary of the nucleus is called the nuclear envelope. It consists of two phospholipid bilayers: an outer membrane and an inner membrane. The nuclear membrane is continuous with the endoplasmic reticulum. Nuclear pores allow substances to enter and exit the nucleus.

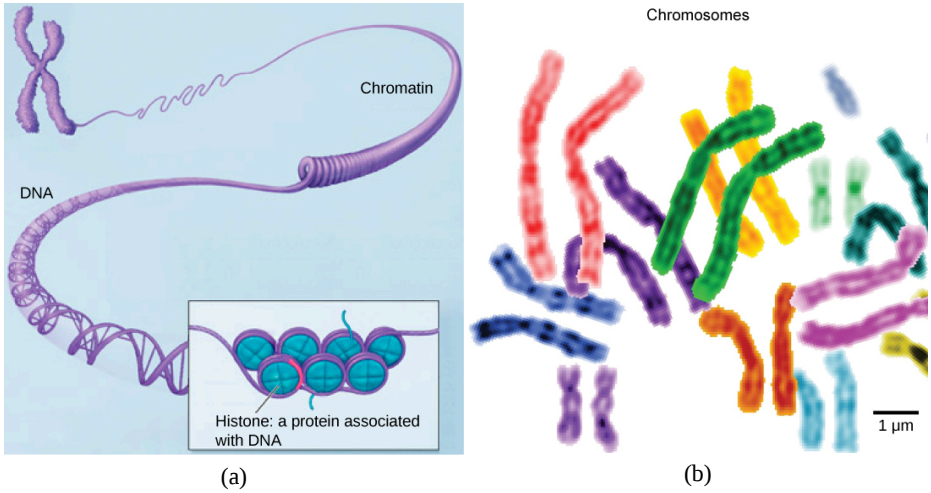
The Nuclear Envelope

The **nuclear envelope** is a double-membrane structure that constitutes the outermost portion of the nucleus (**Figure 9.11**). Both the inner and outer membranes of the nuclear envelope are phospholipid bilayers.

The nuclear envelope is punctuated with pores that control the passage of ions, molecules, and RNA between the nucleoplasm and cytoplasm. The **nucleoplasm** is the semi-solid fluid inside the nucleus, where we find the chromatin and the nucleolus.

Chromatin and Chromosomes

To understand chromatin, it is helpful to first consider chromosomes. **Chromosomes** are structures within the nucleus that are made up of DNA, the hereditary material. You may remember that in prokaryotes, DNA is organized into a single circular chromosome. In eukaryotes, chromosomes are linear structures. Every eukaryotic species has a specific number of chromosomes in the nuclei of its body's cells. For example, in humans, the chromosome number is 46, while in fruit flies, it is eight. Chromosomes are only visible and distinguishable from one another when the cell is getting ready to divide. When the cell is in the growth and maintenance phases of its life cycle, proteins are attached to chromosomes, and they resemble an unwound, jumbled bunch of threads. These unwound protein-chromosome complexes are called **chromatin** (**Figure 9.12**); chromatin describes the material that makes up the chromosomes both when condensed and decondensed.



(a) (b)

Figure 9.12 (a) This image shows various levels of the organization of chromatin (DNA and protein). (b) This image shows paired chromosomes. (credit b: modification of work by NIH; scale-bar data from Matt Russell)

The Nucleolus

We already know that the nucleus directs the synthesis of ribosomes, but how does it do this? Some chromosomes have sections of DNA that encode ribosomal RNA. A darkly staining area within the nucleus called the **nucleolus** (plural = nucleoli) aggregates the ribosomal RNA with associated proteins to assemble the ribosomal subunits that are then transported out through the pores in the nuclear envelope to the cytoplasm.

Ribosomes

Ribosomes are the cellular organelles responsible for protein synthesis. When viewed through an electron microscope, ribosomes appear either as clusters (polyribosomes) or single, tiny dots that float freely in the cytoplasm. They may be attached to the cytoplasmic side of the plasma membrane or the cytoplasmic side of the endoplasmic reticulum and the outer membrane of the nuclear envelope (**Figure 9.8**). Electron microscopy has shown us that ribosomes, which are large complexes of protein and RNA, consist of two subunits, aptly called large and small (**Figure 9.13**). Ribosomes receive their "orders" for protein synthesis from the nucleus where the DNA is transcribed into messenger RNA (mRNA). The mRNA travels to the ribosomes, which translate the code provided by the sequence of the nitrogenous bases in the mRNA into a specific order of amino acids in a protein. Amino acids are the building blocks of proteins.

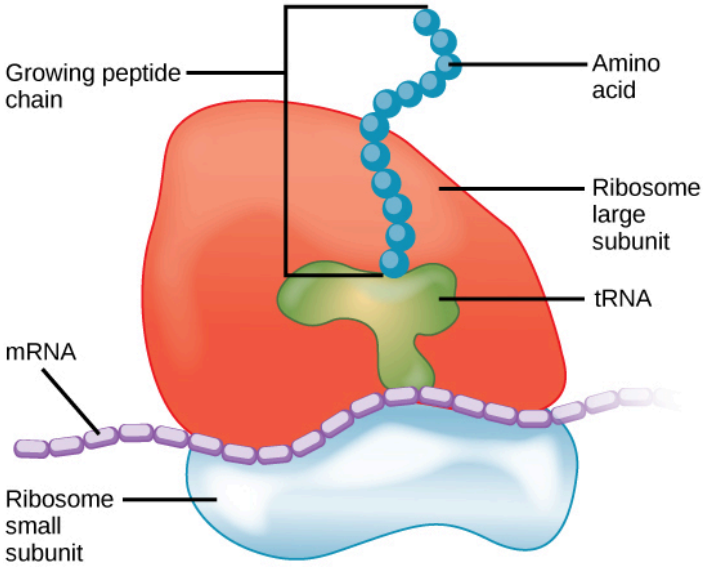


Figure 9.13 Ribosomes are made up of a large subunit (top) and a small subunit (bottom). During protein synthesis, ribosomes assemble amino acids into proteins.

Because proteins synthesis is an essential function of all cells (including enzymes, hormones, antibodies, pigments, structural components, and surface receptors), ribosomes are found in practically every cell. Ribosomes are particularly abundant in cells that synthesize large amounts of protein. For example, the pancreas is responsible for creating several digestive enzymes and the cells that produce these enzymes contain many ribosomes. Thus, we see another example of form following function.

Mitochondria

Mitochondria (singular = mitochondrion) are often called the "powerhouses" or "energy factories" of a cell because they are responsible for making adenosine triphosphate (ATP), the cell's main energy-carrying molecule. ATP represents the short-term stored energy of the cell. Cellular respiration is the process of making ATP using the chemical energy found in glucose and other nutrients. In mitochondria, this process uses oxygen and produces carbon dioxide as a waste product. In fact, the carbon dioxide that you exhale with every breath comes from the cellular reactions that produce carbon dioxide as a byproduct.

In keeping with our theme of form following function, it is important to point out that muscle cells have a very high concentration of mitochondria that produce ATP. Your muscle cells need a lot of energy to keep your body moving. When your cells don't get enough oxygen, they do not make a lot of ATP. Instead, the small amount of ATP they make in the absence of oxygen is accompanied by the production of lactic acid.

Mitochondria are oval-shaped, double membrane organelles (**Figure 9.14**) that have their own ribosomes and DNA. Each membrane is a phospholipid bilayer embedded with proteins. The inner layer has folds called cristae. The area surrounded

by the folds is called the mitochondrial matrix. The cristae and the matrix have different roles in cellular respiration.

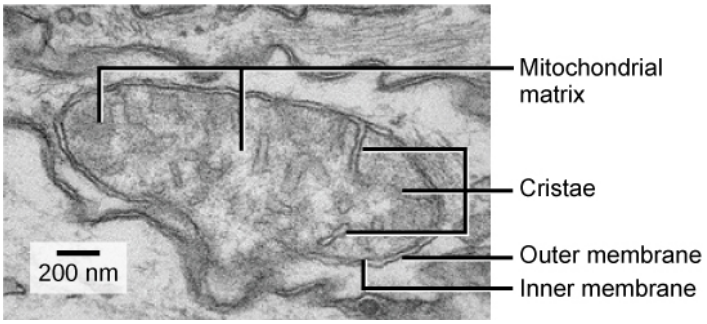


Figure 9.14 This electron micrograph shows a mitochondrion as viewed with a transmission electron microscope. This organelle has an outer membrane and an inner membrane. The inner membrane contains folds, called cristae, which increase its surface area. The space between the two membranes is called the intermembrane space, and the space inside the inner membrane is called the mitochondrial matrix. ATP synthesis takes place on the inner membrane. (credit: modification of work by Matthew Britton; scale-bar data from Matt Russell)

Peroxisomes

Peroxisomes are small, round organelles enclosed by single membranes. They carry out oxidation reactions that break down fatty acids and amino acids. They also detoxify many poisons that may enter the body. (Many of these oxidation reactions release hydrogen peroxide, H_2O_2, which would be damaging to cells; however, when these reactions are confined to peroxisomes, enzymes safely break down the H_2O_2 into oxygen and water.) For example, alcohol is detoxified by peroxisomes in liver cells. Glyoxysomes, which are specialized peroxisomes in plants, are responsible for converting stored fats into sugars.

Vesicles and Endosomes

Vesicles and **endosomes** are membrane-bound sacs that function in storage and transport. Other than the fact that vacuoles are somewhat larger than vesicles, there is a very subtle distinction between them: The membranes of vesicles can fuse with either the plasma membrane or other membrane systems within the cell. Additionally, some agents such as enzymes within plant vacuoles break down macromolecules. The membrane of a vacuole does not fuse with the membranes of other cellular components.

Animal Cells versus Plant Cells

At this point, you know that each eukaryotic cell has a plasma membrane, cytoplasm, a nucleus, ribosomes, mitochondria, peroxisomes, and in some, vacuoles, but there are some striking differences between animal and plant cells. While both animal and plant cells have microtubule organizing centers (MTOCs), animal cells also have centrioles associated with the MTOC: a complex called the centrosome. Animal cells each have a centrosome and lysosomes, whereas plant cells do not. Plant cells have a cell wall, chloroplasts and other specialized plastids, and a large central vacuole, whereas animal cells do not.

The Centrosome

The **centrosome** is a microtubule-organizing center found near the nuclei of animal cells. It contains a pair of centrioles, two structures that lie perpendicular to each other (**Figure 9.15**). Each centriole is a cylinder of nine triplets of microtubules.

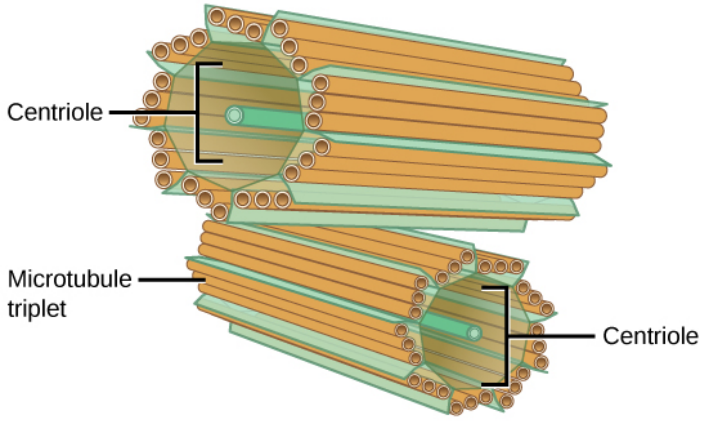


Figure 9.15 The centrosome consists of two centrioles that lie at right angles to each other. Each centriole is a cylinder made up of nine triplets of microtubules. Nontubulin proteins (indicated by the green lines) hold the microtubule triplets together.

The centrosome (the organelle where all microtubules originate) replicates itself before a cell divides, and the centrioles appear to have some role in pulling the duplicated chromosomes to opposite ends of the dividing cell. However, the exact function of the centrioles in cell division isn't clear, because cells that have had the centrosome removed can still divide, and plant cells, which lack centrosomes, are capable of cell division.

Lysosomes

Another set of organelles only found in eukaryotes are lysosomes. The **lysosomes** are the cell's "garbage disposal." Enzymes within the lysosomes aid the breakdown of proteins, polysaccharides, lipids, nucleic acids, and even worn-out organelles. These enzymes are active at a much lower pH than that of the cytoplasm. Therefore, the pH within lysosomes is more acidic than the pH of the cytoplasm. Many reactions that take place in the cytoplasm could not occur at a low pH, so again, the advantage of compartmentalizing the eukaryotic cell into organelles is apparent.

The Cell Wall

If you examine **Figure 9.8b**, the diagram of a plant cell, you will see a structure external to the plasma membrane called the cell wall. The **cell wall** is a rigid covering that protects the cell, provides structural support, and gives shape to the cell. Fungal and protistan cells also have cell walls. While the chief component of prokaryotic cell walls is peptidoglycan, the major organic molecule in the plant cell wall is cellulose (**Figure 9.16**), a polysaccharide made up of glucose units. Have you ever noticed that when you bite into a raw vegetable, like celery, it crunches? That's because you are tearing the rigid cell walls of the celery cells with your teeth.

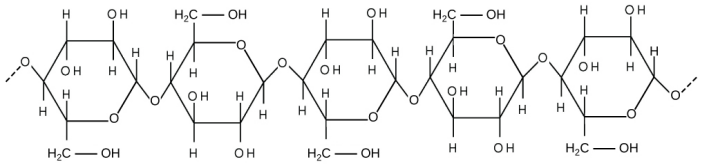


Figure 9.16 Cellulose is a long chain of β-glucose molecules connected by a 1-4 linkage. The dashed lines at each end of the figure indicate a series of many more glucose units. The size of the page makes it impossible to portray an entire cellulose molecule.

Chloroplasts

Like the mitochondria, chloroplasts have their own DNA and ribosomes, but chloroplasts have an entirely different function. **Chloroplasts** are plant cell organelles that carry out photosynthesis. Photosynthesis is the series of reactions that use carbon dioxide, water, and light energy to make glucose and oxygen. This is a major difference between plants and animals; plants (autotrophs) are able to make their own food, like sugars, while animals (heterotrophs) must ingest their food.

Like mitochondria, chloroplasts have outer and inner membranes, but within the space enclosed by a chloroplast's inner membrane is a set of interconnected and stacked fluid-filled membrane sacs called thylakoids (**Figure 9.17**). Each stack of thylakoids is called a granum (plural = grana). The fluid enclosed by the inner membrane that surrounds the grana is called the stroma.

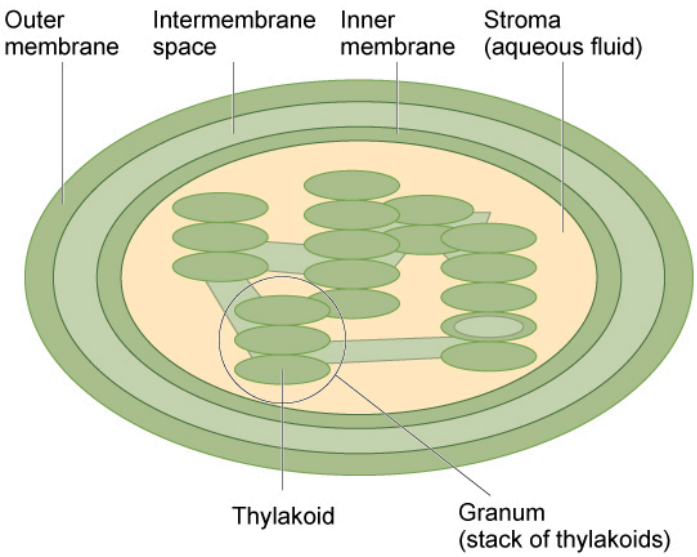


Figure 9.17 The chloroplast has an outer membrane, an inner membrane, and membrane structures called thylakoids that are stacked into grana. The space inside the thylakoid membranes is called the thylakoid space. The light harvesting reactions take place in the thylakoid membranes, and the synthesis of sugar takes place in the fluid inside the inner membrane, which is called the stroma. Chloroplasts also have their own genome, which is contained on a single circular chromosome.

The chloroplasts contain a green pigment called **chlorophyll**, which captures the light energy that drives the reactions of photosynthesis. Like plant cells, photosynthetic protists also have chloroplasts. Some bacteria perform photosynthesis, but their chlorophyll is not relegated to an organelle.

The Central Vacuole

Previously, we mentioned vacuoles as essential components of plant cells. If you look at **Figure 9.8b**, you will see that plant cells each have a large central vacuole that occupies most of the area of the cell. The **central vacuole** plays a key role in regulating the cell's concentration of water in changing environmental conditions. Have you ever noticed that if you forget to water a plant for a few days, it wilts? That's because as the water concentration in the soil becomes lower than the water concentration in the plant, water moves out of the central vacuoles and cytoplasm. As the central vacuole shrinks, it leaves the cell wall unsupported. This loss of support to the cell walls of plant cells results in the wilted appearance of the plant.

The central vacuole also supports the expansion of the cell. When the central vacuole holds more water, the cell gets larger without having to invest a lot of energy in synthesizing new cytoplasm.

9.4 | Protists

Introduction

> "The 31th of May, I perceived in the same water more of those Animals, as also some that were somewhat bigger. And I imagine that ten hundred thousand of these little Creatures do not equal an ordinary grain of Sand in bigness..."

Antoni von Leeuwenhoek, Letter to H. Oldenburg, 9 October 1676

Von Leeuwenhoek's amazement at seeing protists for the first time is understandable. These tiny creatures are abundant, diverse, and fit into many biological niches. There are over 100,000 described living species of protists, and it is unclear how many undescribed species may exist. Since many protists live as commensals or parasites in other organisms and these relationships are often species-specific, there is a huge potential for protist diversity that matches the diversity of hosts. As

the catchall term for eukaryotic organisms that are not animal, plant, or fungi, it is not surprising that very few characteristics are common to all protists. Since Protista is a catchall group of organisms, Biologists are now investigating the evolutionary relationships of this groups and are formulating many new Kingdoms within the group of Protists.

Cell Structure

The cells of protists are among the most elaborate of all cells. Most protists are microscopic and unicellular, but some true multicellular forms exist. A few protists live as colonies that behave in some ways as a group of free-living cells and in other ways as a multicellular organism. Still other protists are composed of enormous, multinucleate, single cells that look like amorphous blobs of slime, or in other cases, like ferns. In fact, many protist cells are multinucleated; in some species, the nuclei are different sizes and have distinct roles in protist cell function.

Single protist cells range in size from less than a micrometer to three meters in length to hectares. Protist cells may be enveloped by animal-like plasma membranes or plant-like cell walls. Others are encased in glassy silica-based shells or wound with **pellicles** of interlocking protein strips. The pellicle functions like a flexible coat of armor, preventing the protist from being torn or pierced without compromising its range of motion.

Metabolism

Protists exhibit many forms of nutrition and may be aerobic or anaerobic. Protists that store energy by photosynthesis belong to a group of photoautotrophs and are characterized by the presence of chloroplasts. Other protists are heterotrophic and consume organic materials (such as other organisms) to obtain nutrition. Amoebas and some other heterotrophic protist species ingest particles by a process called phagocytosis, in which the plasma membrane engulfs a food particle and brings it inward, pinching off an intracellular membranous sac, or vesicle, called a food vacuole (Figure 9.18). The vesicle containing the ingested particle, the phagosome, then fuses with a lysosome containing hydrolytic enzymes to produce a **phagolysosome**, and the food particle is broken down into small molecules that can diffuse into the cytoplasm and be used in cellular metabolism. Undigested remains ultimately are expelled from the cell via exocytosis.

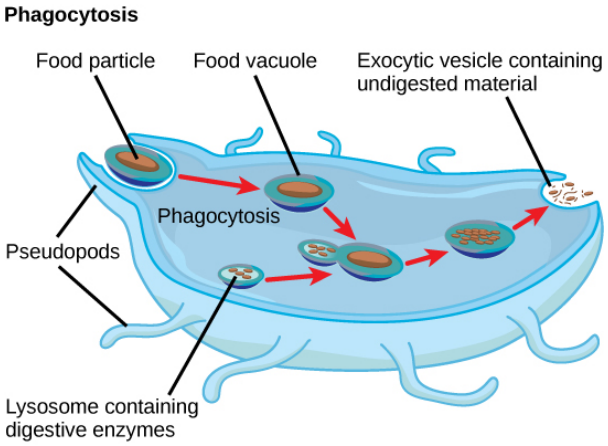


Figure 9.18 The stages of phagocytosis include the engulfment of a food particle, the digestion of the particle using hydrolytic enzymes contained within a lysosome, and the expulsion of undigested materials from the cell.

Subtypes of heterotrophs, called saprobes, absorb nutrients from dead organisms or their organic wastes. Some protists can function as **mixotrophs**, obtaining nutrition by photoautotrophic or heterotrophic routes, depending on whether sunlight or organic nutrients are available.

Motility

The majority of protists are motile, but different types of protists have evolved varied modes of movement (Figure 9.19). Some protists have one or more flagella, which they rotate or whip. Others are covered in rows or tufts of tiny cilia that they coordinately beat to swim. Still others form cytoplasmic extensions called pseudopodia anywhere on the cell, anchor the pseudopodia to a substrate, and pull themselves forward. Some protists can move toward or away from a stimulus, a movement referred to as taxis. Movement toward light, termed phototaxis, is accomplished by coupling their locomotion strategy with a light-sensing organ.

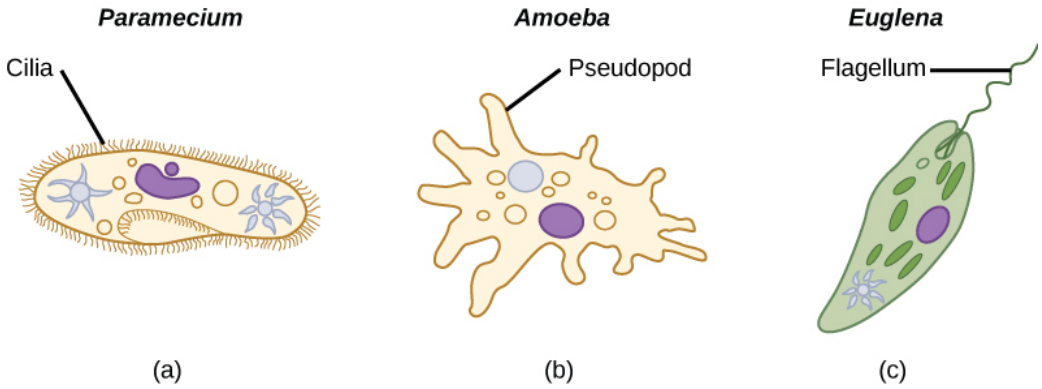


Figure 9.19 Protists use various methods for transportation. (a) *Paramecium* waves hair-like appendages called cilia to propel itself. (b) *Amoeba* uses lobe-like pseudopodia to anchor itself to a solid surface and pull itself forward. (c) *Euglena* uses a whip-like tail called a flagellum to propel itself.

Life Cycles

Protists reproduce by a variety of mechanisms. Most undergo some form of asexual reproduction, such as binary fission, to produce two daughter cells. In protists, binary fission can be divided into transverse or longitudinal, depending on the axis of orientation; sometimes *Paramecium* exhibits this method. Some protists such as the true slime molds exhibit multiple fission and simultaneously divide into many daughter cells. Others produce tiny buds that go on to divide and grow to the size of the parental protist. Sexual reproduction, involving meiosis and fertilization, is common among protists, and many protist species can switch from asexual to sexual reproduction when necessary. Sexual reproduction is often associated with periods when nutrients are depleted or environmental changes occur. Sexual reproduction may allow the protist to recombine genes and produce new variations of progeny that may be better suited to surviving in the new environment. However, sexual reproduction is often associated with resistant cysts that are a protective, resting stage. Depending on their habitat, the cysts may be particularly resistant to temperature extremes, desiccation, or low pH. This strategy also allows certain protists to “wait out” stressors until their environment becomes more favorable for survival or until they are carried (such as by wind, water, or transport on a larger organism) to a different environment, because cysts exhibit virtually no cellular metabolism.

Protist life cycles range from simple to extremely elaborate. Certain parasitic protists have complicated life cycles and must infect different host species at different developmental stages to complete their life cycle. Some protists are unicellular in the haploid form and multicellular in the diploid form, a strategy employed by animals. Other protists have multicellular stages in both haploid and diploid forms, a strategy called alternation of generations that is also used by plants.

Habitats

Nearly all protists exist in some type of aquatic environment, including freshwater and marine environments, damp soil, and even snow. Several protist species are parasites that infect animals or plants. A few protist species live on dead organisms or their wastes, and contribute to their decay.

9.5 | Fungi

Introduction

> “When brewer's yeast is mixed with water the microscope reveals that the yeast dissolves into endless small balls, which are scarcely 1/800th of a line in diameter... If these small balls are placed in sugar water, it can be seen that they consist of the eggs of animals. As they expand, they burst, and from them develop small creatures that multiply with unbelievable rapidity in a most unheard-of way.”

Friedrich Wöhler, *Annalen der Pharmacie und Chemie*, 29:100-104, 1839

Wöhler's fanciful depiction of yeast as tiny animals, while not reflecting current taxonomic opinions, is understandable. Yeast, a single-celled member of the Kingdom Fungi, share many characteristics with animals, include being heterotrophic. But they also have cell walls, a characteristic that they share with plants. But even though a superficial glance might indicate that Fungi occupy the middle ground between animals and plants, they are unique in several ways, and have many interesting and useful metabolites and products (such as penicillin, or ethanol). So let's briefly introduce the Fungi here, and return to them in a later module for a more detailed look at these mysterious creatures.

(a) (b) (c)

Figure 9.20 The (a) familiar mushroom is only one type of fungus. The brightly colored fruiting bodies of this (b) coral fungus are displayed. This (c) electron micrograph shows the spore-bearing structures of *Aspergillus*, a type of toxic fungi found mostly in soil and plants. (credit a: modification of work by Chris Wee; credit b: modification of work by Cory Zanker; credit c: modification of work by Janice Haney Carr, Robert Simmons, CDC; scale-bar data from Matt Russell)

The word *fungus* comes from the Latin word for mushroom. Indeed, the familiar mushrooms are fungi, but there are many other types of fungi as well (**Figure 9.20**). The kingdom Fungi includes an enormous variety of living organisms collectively referred to as Eumycota, or true fungi. While scientists have identified about 100,000 species of fungi, this is only a fraction of the over 1 million species likely present on Earth. Edible mushrooms, yeasts, black mold, and *Penicillium notatum* (the producer of the antibiotic penicillin) are all members of the kingdom Fungi, which belongs to the domain Eukarya. As eukaryotes, a typical fungal cell contains a true nucleus and many membrane-bound organelles.

Fungi were once considered plant-like organisms; however, DNA comparisons have shown that fungi are more closely related to animals than plants. Fungi are not capable of photosynthesis: They use complex organic compounds as sources of energy and carbon. Some fungal organisms multiply only asexually, whereas others undergo both asexual reproduction and sexual reproduction. Most fungi produce a large number of spores that are disseminated by the wind. Like bacteria, fungi play an essential role in ecosystems, because they are decomposers and participate in the cycling of nutrients by breaking down organic materials into simple molecules.

Fungi often interact with other organisms, forming mutually beneficial or mutualistic associations. Fungi also cause serious infections in plants and animals. For example, Dutch elm disease is a particularly devastating fungal infection that destroys many native species of elm (*Ulmus* spp.). The fungus infects the vascular system of the tree. It was accidentally introduced to North America in the 1900s and decimated elm trees across the continent. Dutch elm disease is caused by the fungus *Ophiostoma ulmi*. The elm bark beetle acts as a vector and transmits the disease from tree to tree. Many European and Asiatic elms are less susceptible than American elms.

In humans, fungal infections are generally considered challenging to treat because, unlike bacteria, they do not respond to traditional antibiotic therapy since they are also eukaryotes. These infections may prove deadly for individuals with a compromised immune system.

Fungi have many commercial applications. The food industry uses yeasts in baking, brewing, and wine making. Many industrial compounds are byproducts of fungal fermentation. Fungi are the source of many commercial enzymes and antibiotics.

Cell Structure and Function

Fungi are eukaryotes and as such have a complex cellular organization (**Figure 9.210**). As eukaryotes, fungal cells contain a membrane-bound nucleus. A few types of fungi have structures comparable to the plasmids (loops of DNA) seen in bacteria.

Fungal cells also contain mitochondria and a complex system of internal membranes, including the endoplasmic reticulum and Golgi apparatus.

Fungal cells do not have chloroplasts. Although the photosynthetic pigment chlorophyll is absent, many fungi display bright colors, ranging from red to green to black. The poisonous *Amanita muscaria* (fly agaric) is recognizable by its bright red cap with white patches. Other fungi can be red, or purple, or blue, although many are simply white or brownish. Pigments in fungi are associated with the cell wall and play a protective role against ultraviolet radiation. Some pigments are toxic.

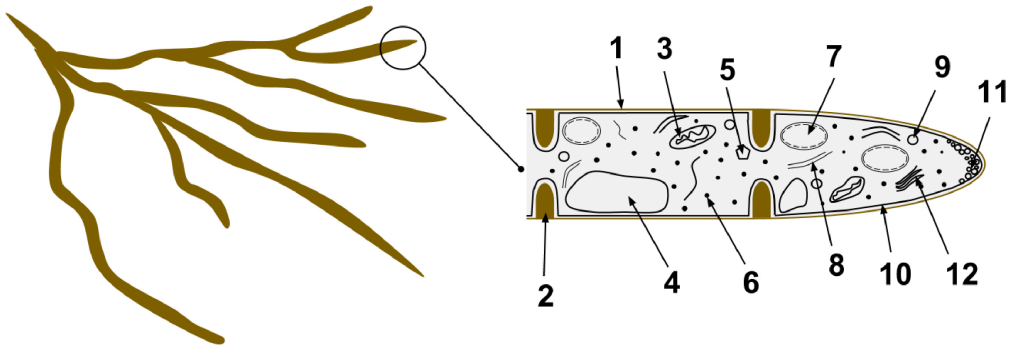


Figure 9.21 Fungal Hyphae. A - Diagram of a branching hyphal mass. B - Diagram of the cellular components of the tip of one of the hyphae. 1- Cell wall 2- Septum 3- Mitochondrion 4- Vacuole 5- Ergosterol crystal 6- Ribosome 7- Nucleus 8- Endoplasmic reticulum 9- Lipid body 10- Plasma membrane 11- Growth tip and vesicles 12- Golgi apparatus. (From Creative Commons, original artist - AHiggins12)

Like plant cells, fungal cells are surrounded by a thick cell wall; however, the rigid layers contain the complex polysaccharides chitin and glucan. Cellulose, the main component of plant cell walls, is found rarely in fungi. Chitin, also found in the exoskeleton of insects, gives structural strength to the cell walls of fungi. The cell wall protects the cell from desiccation and predators. Fungi have plasma membranes similar to other eukaryotes, except that the structure is stabilized by ergosterol, a steroid molecule that functions like the cholesterol found in animal plasma membranes. Most members of the kingdom Fungi are nonmotile. Flagella are produced only by the gametes in the primitive division Chytridiomycota.

9.6 | Eukaryotic Origins

Introduction

> “In most of the animals we think we know best (mammals, reptiles, insects), the genomes that determine limbs, eyes, and nervous systems, for example, are very similar to our own. These animals, like us, are doubly genomic. Even some unicellular beings that do not have eyes, limbs, or nervous systems--such as amoebas and paramecia--contain both nuclear and mitochondrial genomes. Plants and algae have these double genomes as well, plus a third genome, of symbiotic origin. During their evolutionary history, they ingested (but did not digest) photosynthetic blue-green bacteria.

> Therefore, all visible photosynthetic organisms have at least three genomes. But many organisms--such as the protists that inhabit termites--contain within them up to five or more genomes.”

Lynn Margulis and Dorion Sagan, "The Beast with Five Genomes", *Natural History* Vol. 110 Issue 5, p. 38, 2001

Living things fall into three large groups: Archaea, Bacteria, and Eukarya. The first two have prokaryotic cells, and the third contains all eukaryotes. But, as noted above, eukaryotes contain multiple genomes, and those genomes indicate that eukaryotes probably arose by when one ancestor "captured" another. A relatively sparse fossil record is available to help discern what the first members of each of these lineages looked like, so it is possible that all the events that led to the last common ancestor of extant eukaryotes will never be clarified by the fossil record. However, comparative biology of extant organisms, genomic analysis, and the limited fossil record provide some insight into the history of Eukarya.

The earliest fossils found appear to be Bacteria, most likely cyanobacteria. They are about 3.5 billion years old and are recognizable because of their relatively complex structure and, for prokaryotes, relatively large cells. Most other prokaryotes have small cells, 1 or 2 μm in size, and would be difficult to pick out as fossils. Most living eukaryotes have cells measuring 10 μm or greater. Structures this size, which might be fossils, appear in the geological record about 2.1 billion years ago.

Characteristics of Eukaryotes

Data from these fossils have led comparative biologists to the conclusion that living eukaryotes are all descendants of a single common ancestor. Mapping the characteristics found in all major groups of eukaryotes reveals that the following characteristics must have been present in the last common ancestor, because these characteristics are present in at least some of the members of each major lineage.

1. Cells with nuclei surrounded by a nuclear envelope with nuclear pores. This is the single characteristic that is both necessary and sufficient to define an organism as a eukaryote. All extant eukaryotes have cells with nuclei.
2. Mitochondria. Some extant eukaryotes have very reduced remnants of mitochondria in their cells, whereas other members of their lineages have “typical” mitochondria.
3. A cytoskeleton containing the structural and motility components called actin microfilaments and microtubules. All extant eukaryotes have these cytoskeletal elements.
4. Flagella and cilia, organelles associated with cell motility. Some extant eukaryotes lack flagella and/or cilia, but they are descended from ancestors that possessed them.
5. Chromosomes, each consisting of a linear DNA molecule coiled around basic (alkaline) proteins called histones. The few eukaryotes with chromosomes lacking histones clearly evolved from ancestors that had them.
6. Mitosis, a process of nuclear division wherein replicated chromosomes are divided and separated using elements of the cytoskeleton. Mitosis is universally present in eukaryotes.
7. Sex, a process of genetic recombination unique to eukaryotes in which diploid nuclei at one stage of the life cycle undergo meiosis to yield haploid nuclei and subsequent karyogamy, a stage where two haploid nuclei fuse together to create a diploid zygote nucleus.
8. Members of all major lineages have cell walls, and it might be reasonable to conclude that the last common ancestor could make cell walls during some stage of its life cycle. However, not enough is known about eukaryotes’ cell walls and their development to know how much homology exists among them. If the last common ancestor could make cell walls, it is clear that this ability must have been lost in many groups.

Endosymbiosis and the Evolution of Eukaryotes

In order to understand eukaryotic organisms fully, it is necessary to understand that all extant eukaryotes are descendants of a chimeric organism that was a composite of a host cell and the cell(s) of an alpha-proteobacterium that “took up residence” inside it. This major theme in the origin of eukaryotes is known as **endosymbiosis**, one cell engulfing another such that the engulfed cell survives and both cells benefit. Over many generations, a symbiotic relationship can result in two organisms that depend on each other so completely that neither could survive on its own. Endosymbiotic events likely contributed to the origin of the last common ancestor of today’s eukaryotes and to later diversification in certain lineages of eukaryotes (**Figure 9.25**). Before explaining this further, it is necessary to consider metabolism in prokaryotes.

Prokaryotic Metabolism

Many important metabolic processes arose in prokaryotes, and some of these, such as nitrogen fixation, are never found in eukaryotes. The process of aerobic respiration is found in all major lineages of eukaryotes, and it is localized in the mitochondria. Aerobic respiration is also found in many lineages of prokaryotes, but it is not present in all of them, and many forms of evidence suggest that such anaerobic prokaryotes never carried out aerobic respiration nor did their ancestors.

While today's atmosphere is about one-fifth molecular oxygen (O_2), geological evidence shows that it originally lacked O_2. Without oxygen, aerobic respiration would not be expected, and living things would have relied on fermentation instead. At some point before, about 3.5 billion years ago, some prokaryotes began using energy from sunlight to power anabolic processes that reduce carbon dioxide to form organic compounds. That is, they evolved the ability to photosynthesize. Hydrogen, derived from various sources, was captured using light-powered reactions to reduce fixed carbon dioxide in the Calvin cycle. The group of Gram-negative bacteria that gave rise to cyanobacteria used water as the hydrogen source and released O_2 as a waste product.

Eventually, the amount of photosynthetic oxygen built up in some environments to levels that posed a risk to living organisms, since it can damage many organic compounds. Various metabolic processes evolved that protected organisms from oxygen, one of which, aerobic respiration, also generated high levels of ATP. It became widely present among prokaryotes, including in a group we now call alpha-proteobacteria. Organisms that did not acquire aerobic respiration had to remain in oxygen-free environments. Originally, oxygen-rich environments were likely localized around places where cyanobacteria were active, but by about 2 billion years ago, geological evidence shows that oxygen was building up to higher concentrations in the atmosphere. Oxygen levels similar to today's levels only arose within the last 700 million years.

Recall that the first fossils that we believe to be eukaryotes date to about 2 billion years old, so they appeared as oxygen levels were increasing. Also, recall that all extant eukaryotes descended from an ancestor with mitochondria. These organelles were first observed by light microscopists in the late 1800s, where they appeared to be somewhat worm-shaped structures that seemed to be moving around in the cell. Some early observers suggested that they might be bacteria living inside host cells, but these hypotheses remained unknown or rejected in most scientific communities.

Endosymbiotic Theory

As cell biology developed in the twentieth century, it became clear that mitochondria were the organelles responsible for producing ATP using aerobic respiration. In the 1960s, American biologist Lynn Margulis developed **endosymbiotic theory**, which states that eukaryotes may have been a product of one cell engulfing another, one living within another, and evolving over time until the separate cells were no longer recognizable as such. In 1967, Margulis introduced new work on the theory and substantiated her findings through microbiological evidence. Although Margulis' work initially was met with resistance, this once-revolutionary hypothesis is now widely accepted, with work progressing on uncovering the steps involved in this evolutionary process and the key players involved. Much still remains to be discovered about the origins of the cells that now make up the cells in all living eukaryotes.

Broadly, it has become clear that many of our nuclear genes and the molecular machinery responsible for replication and expression appear closely related to those in Archaea. On the other hand, the metabolic organelles and genes responsible for many energy-harvesting processes had their origins in bacteria. Much remains to be clarified about how this relationship occurred; this continues to be an exciting field of discovery in biology. For instance, it is not known whether the endosymbiotic event that led to mitochondria occurred before or after the host cell had a nucleus. Such organisms would be among the extinct precursors of the last common ancestor of eukaryotes.

Mitochondria

One of the major features distinguishing prokaryotes from eukaryotes is the presence of **mitochondria**. Eukaryotic cells may contain anywhere from one to several thousand mitochondria, depending on the cell's level of energy consumption. Each mitochondrion measures 1 to 10 or greater micrometers in length and exists in the cell as an organelle that can be ovoid to worm-shaped to intricately branched (**Figure 9.22**). Mitochondria arise from the division of existing mitochondria; they may fuse together; and they may be moved around inside the cell by interactions with the cytoskeleton. However, mitochondria cannot survive outside the cell. As the atmosphere was oxygenated by photosynthesis, and as successful aerobic prokaryotes evolved, evidence suggests that an ancestral cell with some membrane compartmentalization engulfed a free-living aerobic prokaryote, specifically an alpha-proteobacterium, thereby giving the host cell the ability to use oxygen to release energy stored in nutrients. Alpha-proteobacteria are a large group of bacteria that includes species symbiotic with plants, disease organisms that can infect humans via ticks, and many free-living species that use light for energy. Several lines of evidence support that mitochondria are derived from this endosymbiotic event. Most mitochondria are shaped like alpha-proteobacteria and are surrounded by two membranes, which would result when one membrane-bound organism was engulfed into a vacuole by another membrane-bound organism. The mitochondrial inner membrane is extensive and involves substantial infoldings called **cristae** that resemble the textured, outer surface of alpha-proteobacteria. The **matrix**

and inner membrane are rich with the enzymes necessary for aerobic respiration.

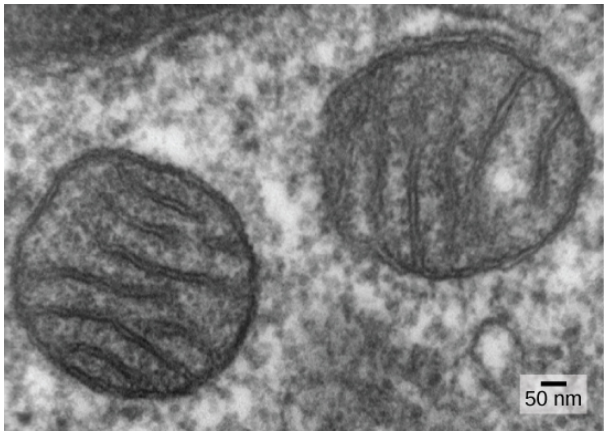


Figure 9.22 In this transmission electron micrograph of mitochondria in a mammalian lung cell, the cristae, infoldings of the mitochondrial inner membrane, can be seen in cross-section. (credit: Louise Howard)

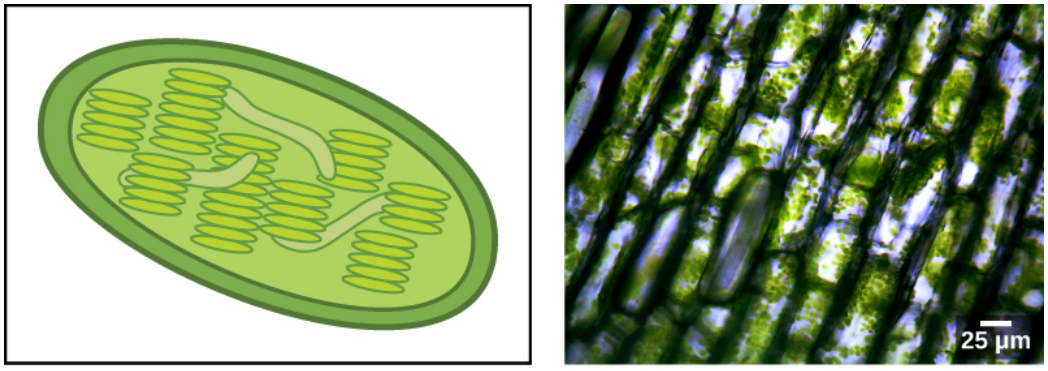
Mitochondria divide independently by a process that resembles binary fission in prokaryotes. Specifically, mitochondria are not formed from scratch (de novo) by the eukaryotic cell; they reproduce within it and are distributed with the cytoplasm when a cell divides or two cells fuse. Therefore, although these organelles are highly integrated into the eukaryotic cell, they still reproduce as if they are independent organisms within the cell. However, their reproduction is synchronized with the activity and division of the cell. Mitochondria have their own (usually) circular DNA chromosome that is stabilized by attachments to the inner membrane and carries genes similar to genes expressed by alpha-proteobacteria. Mitochondria also have special ribosomes and transfer RNAs that resemble these components in prokaryotes. These features all support that mitochondria were once free-living prokaryotes.

Mitochondria that carry out aerobic respiration have their own genomes, with genes similar to those in alpha-proteobacteria. However, many of the genes for respiratory proteins are located in the nucleus. When these genes are compared to those of other organisms, they appear to be of alpha-proteobacterial origin. Additionally, in some eukaryotic groups, such genes are found in the mitochondria, whereas in other groups, they are found in the nucleus. This has been interpreted as evidence that genes have been transferred from the endosymbiont chromosome to the host genome. This loss of genes by the endosymbiont is probably one explanation why mitochondria cannot live without a host.

Some living eukaryotes are anaerobic and cannot survive in the presence of too much oxygen. Some appear to lack organelles that could be recognized as mitochondria. In the 1970s to the early 1990s, many biologists suggested that some of these eukaryotes were descended from ancestors whose lineages had diverged from the lineage of mitochondrion-containing eukaryotes before endosymbiosis occurred. However, later findings suggest that reduced organelles are found in most, if not all, anaerobic eukaryotes, and that all eukaryotes appear to carry some genes in their nuclei that are of mitochondrial origin. In addition to the aerobic generation of ATP, mitochondria have several other metabolic functions. One of these functions is to generate clusters of iron and sulfur that are important cofactors of many enzymes. Such functions are often associated with the reduced mitochondrion-derived organelles of anaerobic eukaryotes. Therefore, most biologists accept that the last common ancestor of eukaryotes had mitochondria.

Plastids

Some groups of eukaryotes are photosynthetic. Their cells contain, in addition to the standard eukaryotic organelles, another kind of organelle called a **plastid**. When such cells are carrying out photosynthesis, their plastids are rich in the pigment chlorophyll *a* and a range of other pigments, called accessory pigments, which are involved in harvesting energy from light. Photosynthetic plastids are called chloroplasts (**Figure 9.23**).



(a) (b)

Figure 9.23 (a) This chloroplast cross-section illustrates its elaborate inner membrane organization. Stacks of thylakoid membranes compartmentalize photosynthetic enzymes and provide scaffolding for chloroplast DNA. (b) In this micrograph of *Elodea* sp., the chloroplasts can be seen as small green spheres. (credit b: modification of work by Brandon Zierer; scale-bar data from Matt Russell)

Like mitochondria, plastids appear to have an endosymbiotic origin. This hypothesis was also championed by Lynn Margulis. Plastids are derived from cyanobacteria that lived inside the cells of an ancestral, aerobic, heterotrophic eukaryote. This is called primary endosymbiosis, and plastids of primary origin are surrounded by two membranes. The best evidence is that this has happened twice in the history of eukaryotes. In one case, the common ancestor of the major lineage/supergroup Archaeplastida took on a cyanobacterial endosymbiont; in the other, the ancestor of the small amoeboid rhizarian taxon, *Paulinella*, took on a different cyanobacterial endosymbiont. Almost all photosynthetic eukaryotes are descended from the first event, and only a couple of species are derived from the other.

Cyanobacteria are a group of Gram-negative bacteria with all the conventional structures of the group. However, unlike most prokaryotes, they have extensive, internal membrane-bound sacs called thylakoids. Chlorophyll is a component of these membranes, as are many of the proteins of the light reactions of photosynthesis. Cyanobacteria also have the peptidoglycan wall and lipopolysaccharide layer associated with Gram-negative bacteria.

Chloroplasts of primary origin have thylakoids, a circular DNA chromosome, and ribosomes similar to those of cyanobacteria. Each chloroplast is surrounded by two membranes. In the group of Archaeplastida called the glaucophytes and in *Paulinella*, a thin peptidoglycan layer is present between the outer and inner plastid membranes. All other plastids lack this relictual cyanobacterial wall. The outer membrane surrounding the plastid is thought to be derived from the vacuole in the host, and the inner membrane is thought to be derived from the plasma membrane of the symbiont.

There is also, as with the case of mitochondria, strong evidence that many of the genes of the endosymbiont were transferred to the nucleus. Plastids, like mitochondria, cannot live independently outside the host. In addition, like mitochondria, plastids are derived from the division of other plastids and never built from scratch. Researchers have suggested that the endosymbiotic event that led to Archaeplastida occurred 1 to 1.5 billion years ago, at least 5 hundred million years after the fossil record suggests that eukaryotes were present.

Not all plastids in eukaryotes are derived directly from primary endosymbiosis. Some of the major groups of algae became photosynthetic by secondary endosymbiosis, that is, by taking in either green algae or red algae (both from Archaeplastida) as endosymbionts (**Figure 9.24ab**). Numerous microscopic and genetic studies have supported this conclusion. Secondary plastids are surrounded by three or more membranes, and some secondary plastids even have clear remnants of the nucleus of endosymbiotic alga. Others have not “kept” any remnants. There are cases where tertiary or higher-order endosymbiotic events are the best explanations for plastids in some eukaryotes.

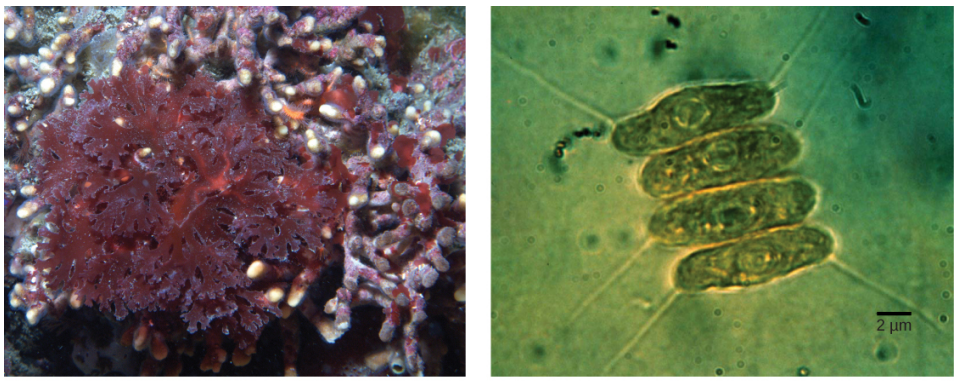


Figure 9.24 (a) Red algae and (b) green algae (visualized by light microscopy) share similar DNA sequences with photosynthetic cyanobacteria. Scientists speculate that, in a process called endosymbiosis, an ancestral prokaryote engulfed a photosynthetic cyanobacterium that evolved into modern-day chloroplasts. (credit a: modification of work by Ed Bierman; credit b: modification of work by G. Fahnenstiel, NOAA; scale-bar data from Matt Russell)

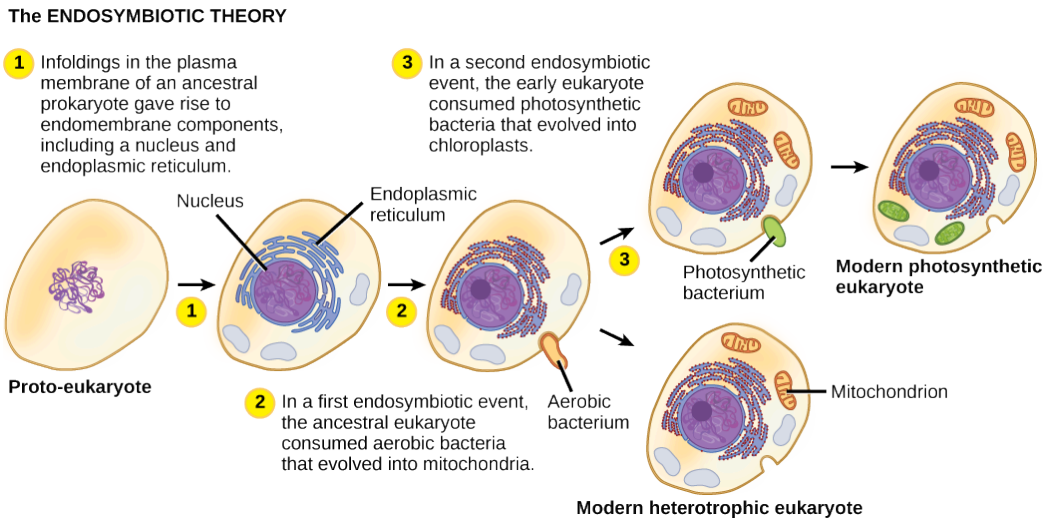


Figure 9.25 The first eukaryote may have originated from an ancestral prokaryote that had undergone membrane proliferation, compartmentalization of cellular function (into a nucleus, lysosomes, and an endoplasmic reticulum), and the establishment of endosymbiotic relationships with an aerobic prokaryote, and, in some cases, a photosynthetic prokaryote, to form mitochondria and chloroplasts, respectively.

evolution CONNECTION

Secondary Endosymbiosis in Chlorarachniophytes

Endosymbiosis involves one cell engulfing another to produce, over time, a coevolved relationship in which neither cell could survive alone. The chloroplasts of red and green algae, for instance, are derived from the engulfment of a photosynthetic cyanobacterium by an early prokaryote.

This leads to the question of the possibility of a cell containing an endosymbiont to itself become engulfed, resulting in a secondary endosymbiosis. Molecular and morphological evidence suggest that the chlorarachniophyte protists are derived from a secondary endosymbiotic event. Chlorarachniophytes are rare algae indigenous to tropical seas and sand that can be classified into the rhizarian supergroup. Chlorarachniophytes extend thin cytoplasmic strands, interconnecting themselves with other chlorarachniophytes, in a cytoplasmic network. These protists are thought to have originated when a eukaryote engulfed a green alga, the latter of which had already established an endosymbiotic relationship with a photosynthetic cyanobacterium (**Figure 9.26**).

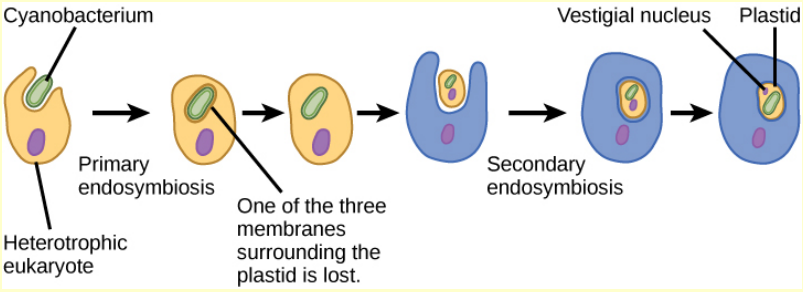


Figure 9.26 The hypothesized process of endosymbiotic events leading to the evolution of chlorarachniophytes is shown. In a primary endosymbiotic event, a heterotrophic eukaryote consumed a cyanobacterium. In a secondary endosymbiotic event, the cell resulting from primary endosymbiosis was consumed by a second cell. The resulting organelle became a plastid in modern chlorarachniophytes.

Several lines of evidence support that chlorarachniophytes evolved from secondary endosymbiosis. The chloroplasts contained within the green algal endosymbionts still are capable of photosynthesis, making chlorarachniophytes photosynthetic. The green algal endosymbiont also exhibits a stunted vestigial nucleus. In fact, it appears that chlorarachniophytes are the products of an evolutionarily recent secondary endosymbiotic event. The plastids of chlorarachniophytes are surrounded by four membranes: The first two correspond to the inner and outer membranes of the photosynthetic cyanobacterium, the third corresponds to the green alga, and the fourth corresponds to the vacuole that surrounded the green alga when it was engulfed by the chlorarachniophyte ancestor. In other lineages that involved secondary endosymbiosis, only three membranes can be identified around plastids. This is currently rectified as a sequential loss of a membrane during the course of evolution.

The process of secondary endosymbiosis is not unique to chlorarachniophytes. In fact, secondary endosymbiosis of green algae also led to euglenid protists, whereas secondary endosymbiosis of red algae led to the evolution of dinoflagellates, apicomplexans, and stramenopiles.

10 | TOUR OF THE CELL: WATER, CARBOHYDRATES AND LIPIDS

10.1 | Atoms, Isotopes, Ions, and Molecules: The Building Blocks

> "Through the discovery of Buchner, Biology was relieved of another fragment of mysticism. The splitting up of sugar into CO_2 and alcohol is no more the effect of a 'vital principle' than the splitting up of cane sugar by invertase. The history of this problem is instructive, as it warns us against considering problems as beyond our reach because they have not yet found their solution."

Jacques Loeb, in *The Dynamics of Living Matter*, (1906)

Loeb is referring to the Nobel Prize-winning experiments of Eduard Buchner, who proved that cells are not necessary for cellular chemical reactions to take place. This was one of the crucial steps toward the synthesis of biology and chemistry that culminates in the modern-day discipline we call biochemistry. The properties of matter are important in the study of biochemistry (and biology), so we need to introduce some chemical concepts to help us with that understanding. At its most fundamental level, life is made up of matter. **Matter** is any substance that occupies space and has mass. **Elements** are unique forms of matter with specific chemical and physical properties that cannot be broken down into smaller substances by ordinary chemical reactions. There are 118 elements, but only 92 occur naturally. The remaining elements have only been synthesized in laboratories, and are unstable.

Each element is designated by its **chemical symbol**, which is a single capital letter or, when the first letter is already "taken" by another element, a combination of two letters. Some elements follow the English term for the element, such as C for carbon and Ca for calcium. Other elements' chemical symbols derive from their Latin names; for example, the symbol for sodium is Na, referring to *natrium*, the Latin word for sodium.

The four most abundant elements in all living organisms are oxygen (O), carbon (C), hydrogen (H), and nitrogen (N). In the non-living world, elements are found in different proportions, and some elements common to living organisms are relatively rare on the earth as a whole, as shown in **Table 10.1**. For example, the atmosphere is rich in nitrogen and oxygen but contains little carbon and hydrogen, while the earth's crust, although it contains oxygen and a small amount of hydrogen, has little nitrogen and carbon. In spite of their differences in abundance, all elements and the chemical reactions between them obey the same chemical and physical laws regardless of whether they are a part of the living or non-living world.

Approximate Percentage of Elements in Living Organisms (Humans) Compared to the Non-living World

Element	Organisms (Humans)	Atmosphere	Earth's Crust
Oxygen (O)	65%	21%	46%
Carbon (C)	18%	trace	trace
Hydrogen (H)	10%	trace	0.1%
Nitrogen (N)	3%	78%	trace

Table 10.1

The Structure of the Atom

To understand how elements come together, we must first discuss the smallest component or building block of an element, the atom. An **atom** is the smallest unit of matter that retains all of the chemical properties of an element. For example, one gold atom has all of the properties of gold in that it is a solid metal at room temperature. A gold coin is simply a very large number of gold atoms molded into the shape of a coin and containing small amounts of other elements known as impurities. Gold atoms cannot be broken down into anything smaller while still retaining the properties of gold.

An atom is composed of two regions: the **nucleus**, which is in the center of the atom and contains **protons** and **neutrons**, and the outermost region of the atom which holds its **electrons** in orbit around the nucleus, as illustrated in **Figure 10.1**. Atoms contain protons, electrons, and neutrons. The only exception is hydrogen (H), which is made of one proton and one electron, with no neutrons.

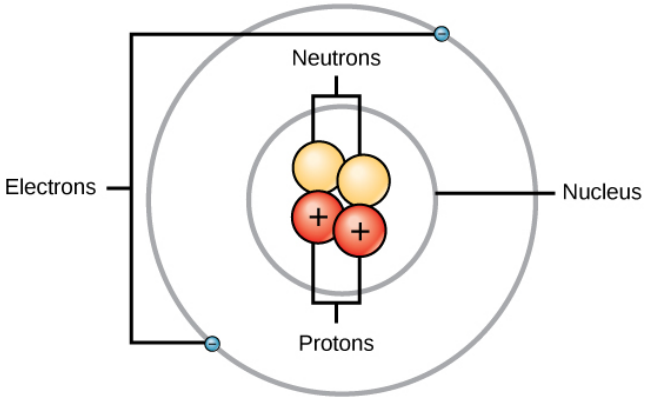


Figure 10.1 Elements, such as helium, depicted here, are made up of atoms. Atoms are made up of protons and neutrons located within the nucleus, with electrons in orbitals surrounding the nucleus.

Protons and neutrons have approximately the same mass, about 1.67×10^{-24} grams. Scientists arbitrarily define this amount of mass as one atomic mass unit (amu) or one Dalton, as shown in **Table 10.2**. Although similar in mass, protons and neutrons differ in their electric charge. A **proton** is positively charged whereas a **neutron** is uncharged. Therefore, the number of neutrons in an atom contributes significantly to its mass, but not to its charge. **Electrons** are much smaller in mass than protons, weighing only 9.11×10^{-28} grams, or about 1/1800 of an atomic mass unit. Hence, they do not contribute much to an element's overall atomic mass. Therefore, when considering atomic mass, it is customary to ignore the mass of any electrons and calculate the atom's mass based on the number of protons and neutrons alone. Although not significant contributors to mass, electrons do contribute greatly to the atom's charge, as each electron has a negative charge equal to the positive charge of a proton. In uncharged, neutral atoms, the number of electrons orbiting the nucleus is equal to the number of protons inside the nucleus. In these atoms, the positive and negative charges cancel each other out, leading to an atom with no net charge.

Accounting for the sizes of protons, neutrons, and electrons, most of the volume of an atom—greater than 99 percent—is, in fact, empty space. With all this empty space, one might ask why so-called solid objects do not just pass through one another. The reason they do not is that the electrons that surround all atoms are negatively charged and negative charges repel each other.

Protons, Neutrons, and Electrons

	Charge	Mass (amu)	Location
Proton	+1	1	nucleus
Neutron	0	1	nucleus
Electron	–1	0	orbitals

Table 10.2

Atomic Number and Mass

Atoms of each element contain a characteristic number of protons and electrons. The number of protons determines an element's **atomic number** and is used to distinguish one element from another. Thus, an atom with only one proton and one electron is always going to be an atom of hydrogen; an atom with two electrons and two protons is an atom of helium, etc. Unlike the fixed number of electrons and protons in atoms of any particular element, however, the number of neutrons is variable. An atom of hydrogen can have zero, one or two neutrons. Different numbers of neutrons will not change the charge, but will change the mass of an atom. These different forms, varying only in the number of neutrons, are called isotopes. Together, the number of protons and the number of neutrons determine an element's mass number, as illustrated in Figure 10.2. Note that the small contribution of mass from electrons is disregarded in calculating the mass number. This approximation of mass can be used to easily calculate how many neutrons an element has by simply subtracting the number of protons from the mass number. Since an element's isotopes will have different mass numbers, scientists also determine the **relative atomic mass** or **standard atomic weight**, which is the calculated mean of the mass numbers for its naturally occurring isotopes. Often, the resulting number contains a fraction. For example, the relative atomic mass of chlorine (Cl) is 35.45 because chlorine is composed of several isotopes, some (the majority) with an atomic mass of 35 (17 protons and 18 neutrons) and some with an atomic mass of 37 (17 protons and 20 neutrons).

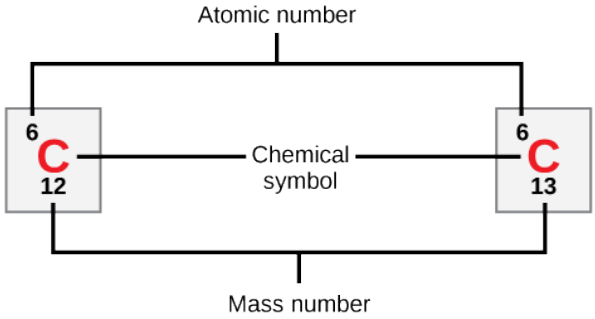


Figure 10.2 Carbon has an atomic number of six, and two stable isotopes with mass numbers of twelve and thirteen, respectively. Its relative atomic mass is 12.011.

Isotopes

Isotopes are different forms of an element that have the same number of protons but a different number of neutrons. Some elements—such as carbon, potassium, and uranium—have naturally occurring isotopes. Carbon-12 contains six protons, six neutrons, and six electrons; therefore, it has a mass number of 12 (six protons and six neutrons). Carbon-14 contains six protons, eight neutrons, and six electrons; its atomic mass is 14 (six protons and eight neutrons). These two alternate forms of carbon are isotopes. Many isotopes are stable, but others are unstable, and can lose neutrons, protons, and electrons, in order to attain a more stable atomic configuration (lower level of potential energy). These unstable isotopes are the radioactive isotopes, or **radioisotopes**. Radioactive decay (carbon-14 losing neutrons to eventually become carbon-12) describes the energy loss that occurs when an unstable atom's nucleus releases particles and radiation.

evolution CONNECTION

Carbon Dating

Carbon is normally present in the atmosphere in the form of gaseous compounds like carbon dioxide and methane. Carbon-14 (^{14}C) is a naturally occurring radioisotope that is created in the atmosphere from atmospheric ^{14}N (nitrogen) by the addition of a neutron and the loss of a proton because of cosmic rays. This is a continuous process, so more ^{14}C is always being created. As a living organism incorporates ^{14}C initially as carbon dioxide fixed in the process of photosynthesis, the relative amount of ^{14}C in its body is equal to the concentration of ^{14}C in the atmosphere. When an organism dies, it is no longer ingesting ^{14}C, so the ratio between ^{14}C and ^{12}C will decline as ^{14}C decays gradually to ^{14}N by a process called beta decay—the emission of electrons or positrons. This decay gives off energy in a slow process.

After approximately 5,730 years, half of the starting concentration of ^{14}C will have been converted back to ^{14}N. The time it takes for half of the original concentration of an isotope to decay back to its more stable form is called its **half-life**. Because the half-life of ^{14}C is long, it is used to date formerly living objects such as old bones or wood. Comparing the ratio of the ^{14}C concentration found in an object to the amount of ^{14}C detected in the atmosphere, the amount of the isotope that has not yet decayed can be determined. On the basis of this amount, the age of the material, such as the pygmy mammoth shown in **Figure 10.3**, can be calculated with accuracy if it is not much older than about 50,000 years. Other elements have isotopes with different half lives. For example, ^{40}K (potassium-40) has a half-life of 1.25 billion years, and ^{235}U (Uranium 235) has a half-life of about 700 million years. Through the use of radiometric dating, scientists can study the age of fossils or other remains of extinct organisms to understand how organisms have evolved from earlier species.

Figure 10.3 The age of carbon-containing remains less than about 50,000 years old, such as this pygmy mammoth, can be determined using carbon dating. (credit: Bill Faulkner, NPS)

The Periodic Table

The different elements are organized and displayed in the periodic table. Devised by Russian chemist Dmitri Mendeleev (1834–1907) in 1869, the table groups elements that, although unique, share certain chemical properties with other elements. The properties of elements are responsible for their physical state at room temperature: they may be gases, solids, or liquids. Elements also have specific chemical reactivity, the ability to combine and to chemically bond with each other.

In the periodic table, shown in **Figure 10.4**, the elements are organized and displayed according to their atomic number and are arranged in a series of rows and columns based on shared chemical and physical properties. In addition to providing the atomic number for each element, the periodic table also displays the element's atomic mass. Looking at carbon, for

example, its symbol (C) and name appear, as well as its atomic number of six (in the upper left-hand corner) and its relative atomic mass of 12.011.

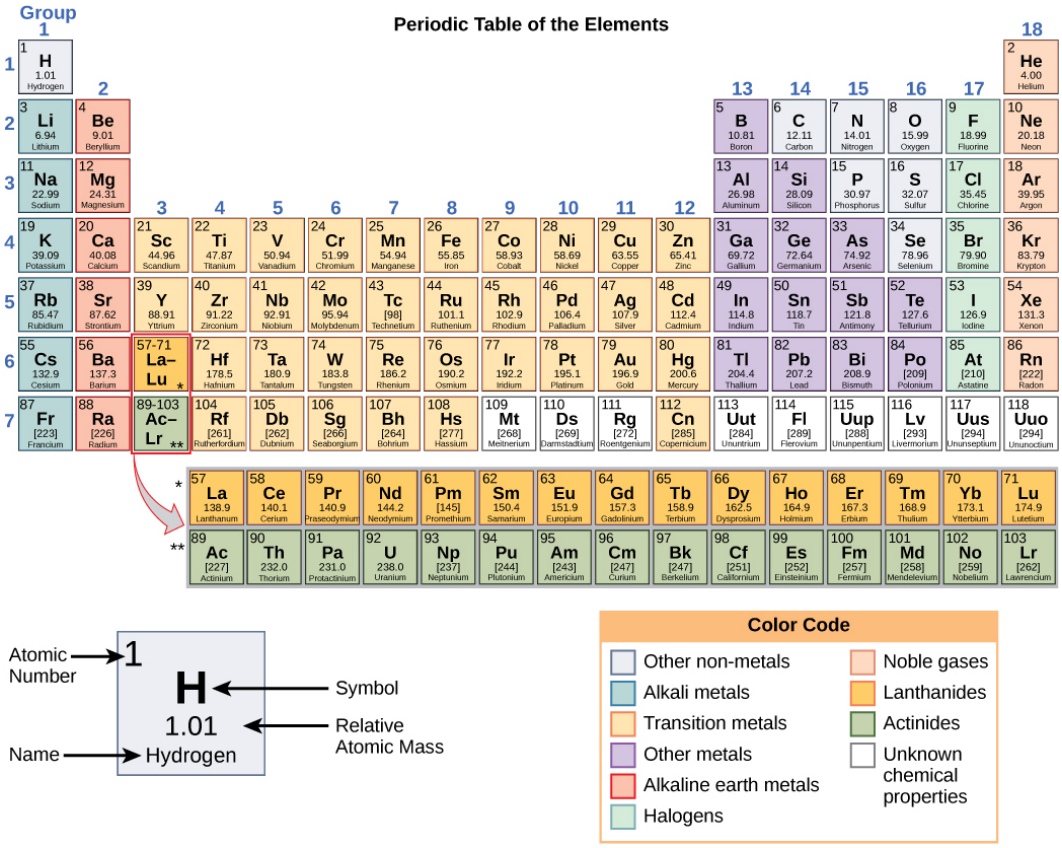


Figure 10.4 The periodic table shows the atomic mass and atomic number of each element. The atomic number appears above the symbol for the element and the approximate atomic mass appears below it.

The periodic table groups elements according to chemical properties. The differences in chemical reactivity between the elements are based on the number and spatial distribution of an atom's electrons. Atoms that chemically react and bond to each other form molecules. **Molecules** are simply two or more atoms chemically bonded together. Logically, when two atoms chemically bond to form a molecule, their electrons, which form the outermost region of each atom, come together first as the atoms form a chemical bond.

Electron Shells and the Bohr Model

It should be stressed that there is a connection between the number of protons in an element, the atomic number that distinguishes one element from another, and the number of electrons it has. In all electrically neutral atoms, the number of electrons is the same as the number of protons. Thus, each element, at least when electrically neutral, has a characteristic number of electrons equal to its atomic number.

An early model of the atom was developed in 1913 by Danish scientist Niels Bohr (1885–1962). The Bohr model shows the atom as a central nucleus containing protons and neutrons, with the electrons in circular orbitals at specific distances from the nucleus, as illustrated in **Figure 10.5**. These orbits form electron shells or energy levels, which are a way of visualizing the number of electrons in the outermost shells. These energy levels are designated by a number and the symbol "n." For example, 1n represents the first energy level located closest to the nucleus.

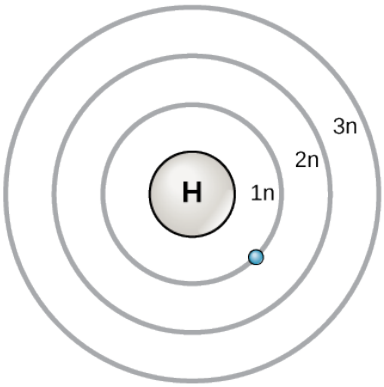


Figure 10.5 The Bohr model was developed by Niels Bohrs in 1913. In this model, electrons exist within principal shells. An electron normally exists in the lowest energy shell available, which is the one closest to the nucleus. Energy from a photon of light can bump it up to a higher energy shell, but this situation is unstable, and the electron quickly decays back to the ground state. In the process, a photon of light is released.

Electrons fill orbitals in a consistent order: they first fill the orbitals closest to the nucleus, then they continue to fill orbitals of increasing energy further from the nucleus. If there are multiple orbitals of equal energy, they will be filled with one electron in each energy level before a second electron is added. The electrons of the outermost energy level determine the energetic stability of the atom and its tendency to form chemical bonds with other atoms to form molecules.

Under standard conditions, atoms fill the inner shells first, often resulting in a variable number of electrons in the outermost shell. The innermost shell has a maximum of two electrons but the next two electron shells can each have a maximum of eight electrons. This is known as the octet rule, which states, with the exception of the innermost shell, that atoms are more stable energetically when they have eight electrons in their valence shell, the outermost electron shell. Examples of some neutral atoms and their electron configurations are shown in **Figure 10.6**. Notice that in this **Figure 10.6**, helium has a complete outer electron shell, with two electrons filling its first and only shell. Similarly, neon has a complete outer 2n shell containing eight electrons. In contrast, chlorine and sodium have seven and one in their outer shells, respectively, but theoretically they would be more energetically stable if they followed the octet rule and had eight.

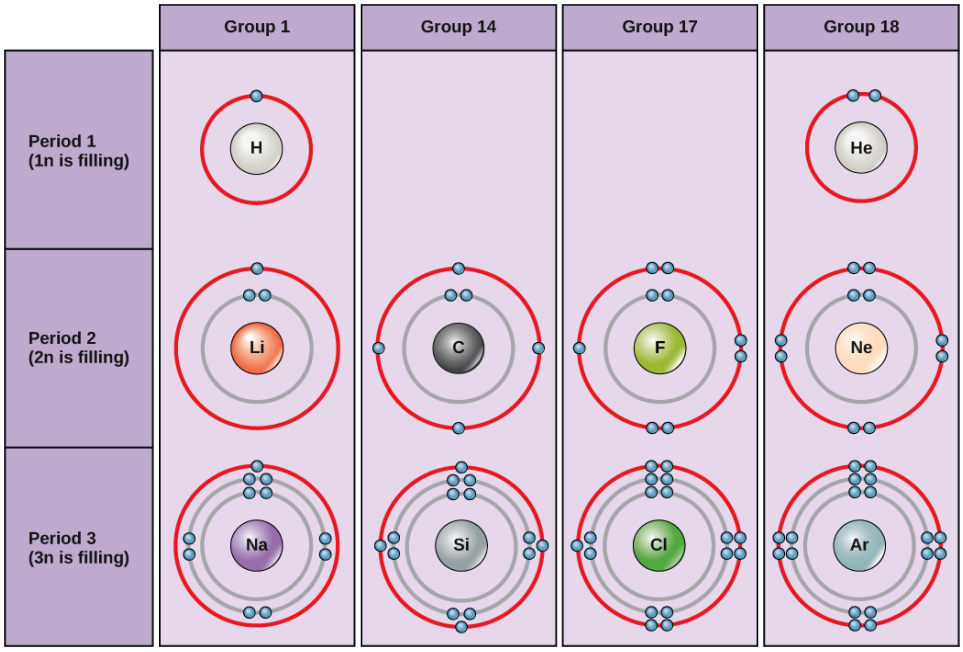


Figure 10.6 Bohr diagrams indicate how many electrons fill each principal shell. Group 18 elements (helium, neon, and argon are shown) have a full outer, or valence, shell. A full valence shell is the most stable electron configuration. Elements in other groups have partially filled valence shells and gain or lose electrons to achieve a stable electron configuration.

Understanding that the organization of the periodic table is based on the total number of protons (and electrons) helps us know how electrons are distributed among the outer shell. The periodic table is arranged in columns and rows based on the number of electrons and where these electrons are located. Take a closer look at the some of the elements in the table's far right column in **Figure 10.4**. The group 18 atoms helium (He), neon (Ne), and argon (Ar) all have filled outer electron shells, making it unnecessary for them to share electrons with other atoms to attain stability; they are highly stable as single atoms. Their non-reactivity has resulted in their being named the inert gases (or noble gases). Compare this to the group 1 elements in the left-hand column. These elements, including hydrogen (H), lithium (Li), and sodium (Na), all have one electron in their outermost shells. That means that they can achieve a stable configuration and a filled outer shell by donating or sharing one electron with another atom or a molecule such as water. Hydrogen will donate or share its electron to achieve this configuration, while lithium and sodium will donate their electron to become stable. As a result of losing a negatively charged electron, they become positively charged ions. Group 17 elements, including fluorine and chlorine, have seven electrons in their outmost shells, so they tend to fill this shell with an electron from other atoms or molecules, making them negatively charged ions. Group 14 elements, of which carbon is the most important to living systems, have four electrons in their outer shell allowing them to make several covalent bonds (discussed below) with other atoms. Thus, the columns of the periodic table represent the potential shared state of these elements' outer electron shells that is responsible for their similar chemical characteristics.

Chemical Reactions and Molecules

All elements are most stable when their outermost shell is filled with electrons according to the octet rule. This is because it is energetically favorable for atoms to be in that configuration and it makes them stable. However, since not all elements have enough electrons to fill their outermost shells, atoms form **chemical bonds** with other atoms thereby obtaining the electrons they need to attain a stable electron configuration. When two or more atoms chemically bond with each other, the resultant chemical structure is a molecule. The familiar water molecule, H_2O, consists of two hydrogen atoms and one oxygen atom; these bond together to form water, as illustrated in **Figure 10.7**. Atoms can form molecules by donating, accepting, or sharing electrons to fill their outer shells.

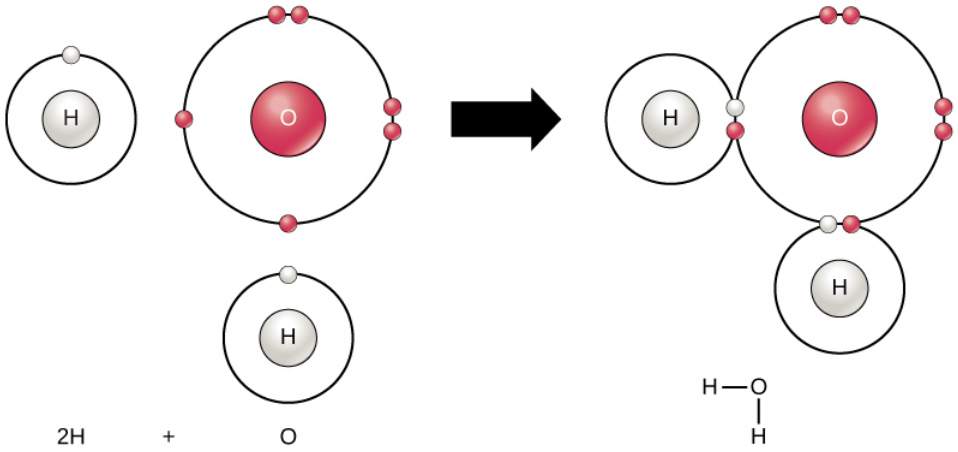


Figure 10.7 Two or more atoms may bond with each other to form a molecule. When two hydrogens and an oxygen share electrons via covalent bonds, a water molecule is formed.

Chemical reactions occur when two or more atoms bond together to form molecules or when bonded atoms are broken apart. The substances used in the beginning of a chemical reaction are called the **reactants** (usually found on the left side of a chemical equation), and the substances found at the end of the reaction are known as the **products** (usually found on the right side of a chemical equation). An arrow is typically drawn between the reactants and products to indicate the direction of the chemical reaction; this direction is not always a "one-way street." For the creation of the water molecule shown above, the chemical equation would be:

$$2H + O \rightarrow H_2O$$

An example of a simple chemical reaction is the breaking down of hydrogen peroxide molecules, each of which consists of two hydrogen atoms bonded to two oxygen atoms (H_2O_2). The reactant hydrogen peroxide is broken down into water, containing one oxygen atom bound to two hydrogen atoms (H_2O), and oxygen, which consists of two bonded oxygen atoms (O_2). In the equation below, the reaction includes two hydrogen peroxide molecules and two water molecules. This is an example of a **balanced chemical equation**, wherein the number of atoms of each element is the same on each side of the equation. According to the law of conservation of matter, the number of atoms before and after a chemical reaction should be equal, such that no atoms are, under normal circumstances, created or destroyed.

$$2H_2O_2 \text{ (hydrogen peroxide)} \rightarrow 2H_2O \text{ (water)} + O_2 \text{ (oxygen)}$$

Even though all of the reactants and products of this reaction are molecules (each atom remains bonded to at least one other atom), in this reaction only hydrogen peroxide and water are representative of a subclass of molecules known as **compounds**: they contain atoms of more than one type of element. Molecular oxygen, on the other hand, as shown in **Figure 10.8**, consists of two doubly bonded oxygen atoms and is not classified as a compound but as an element.

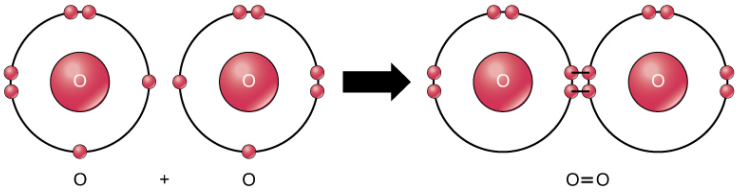


Figure 10.8 The oxygen atoms in an O_2 molecule are joined by a double bond.

Some chemical reactions, such as the one shown above, can proceed in one direction until the reactants are all used up. The equations that describe these reactions contain a unidirectional arrow and are irreversible. Reversible reactions are those that can go in either direction. In reversible reactions, reactants are turned into products, but when the concentration of product goes beyond a certain threshold (characteristic of the particular reaction), some of these products will be converted back into reactants; at this point, the designations of products and reactants are reversed. This back and forth continues until a certain relative balance between reactants and products occurs—a state called equilibrium. These situations of reversible reactions are often denoted by a chemical equation with a double headed arrow pointing towards both the reactants and products.

For example, in human blood, excess hydrogen ions (H^+) bind to bicarbonate ions (HCO_3^-) forming an equilibrium state with carbonic acid (H_2CO_3). If carbonic acid were added to this system, some of it would be converted to bicarbonate and

hydrogen ions.

$$HCO_3^- + H^+ \leftrightarrow H_2CO_3$$

In biological reactions, however, equilibrium is rarely obtained because the concentrations of the reactants or products or both are constantly changing, often with a product of one reaction being a reactant for another. To return to the example of excess hydrogen ions in the blood, the formation of carbonic acid will be the major direction of the reaction. However, the carbonic acid can also leave the body as carbon dioxide gas (via exhalation) instead of being converted back to bicarbonate ion, thus driving the reaction to the right by the chemical law known as law of mass action. These reactions are important for maintaining the homeostasis of our blood.

$$HCO_3^- + H^+ \leftrightarrow H_2CO_3 \leftrightarrow CO_2 + H_2O$$

Ions and Ionic Bonds

Some atoms are more stable when they gain or lose an electron (or possibly two) and form ions. This fills their outermost electron shell and makes them energetically more stable. Because the number of electrons does not equal the number of protons, each ion has a net charge. Cations are positive ions that are formed by losing electrons. Negative ions are formed by gaining electrons and are called anions. Anions are designated by their elemental name being altered to end in "-ide": the anion of chlorine is called chloride, and the anion of sulfur is called sulfide, for example.

This movement of electrons from one element to another is referred to as electron transfer. As **Figure 10.9** illustrates, sodium (Na) only has one electron in its outer electron shell. It takes less energy for sodium to donate that one electron than it does to accept seven more electrons to fill the outer shell. If sodium loses an electron, it now has 11 protons, 11 neutrons, and only 10 electrons, leaving it with an overall charge of +1. It is now referred to as a sodium ion. Chlorine (Cl) in its lowest energy state (called the ground state) has seven electrons in its outer shell. Again, it is more energy-efficient for chlorine to gain one electron than to lose seven. Therefore, it tends to gain an electron to create an ion with 17 protons, 17 neutrons, and 18 electrons, giving it a net negative (–1) charge. It is now referred to as a chloride ion. In this example, sodium will donate its one electron to empty its shell, and chlorine will accept that electron to fill its shell. Both ions now satisfy the octet rule and have complete outermost shells. Because the number of electrons is no longer equal to the number of protons, each is now an ion and has a +1 (sodium cation) or –1 (chloride anion) charge. Note that these transactions can normally only take place simultaneously: in order for a sodium atom to lose an electron, it must be in the presence of a suitable recipient like a chlorine atom.

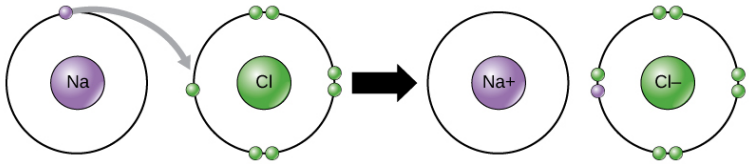


Figure 10.9 In the formation of an ionic compound, metals lose electrons and nonmetals gain electrons to achieve an octet.

Ionic bonds are formed between ions with opposite charges. For instance, positively charged sodium ions and negatively charged chloride ions bond together to make crystals of sodium chloride, or table salt, creating a crystalline molecule with zero net charge.

Certain salts are referred to in physiology as electrolytes (including sodium, potassium, and calcium), ions necessary for nerve impulse conduction, muscle contractions and water balance. Many sports drinks and dietary supplements provide these ions to replace those lost from the body via sweating during exercise.

Covalent Bonds and Other Bonds and Interactions

Another way the octet rule can be satisfied is by the sharing of electrons between atoms to form **covalent bonds**. Covalent bonds are much more common than ionic bonds in the molecules of living organisms, and often the covalent bonds discussed in these systems are stronger than the ionic bonds. So, Biologists often think of covalent bonds as being stronger than ionic bonds, in fact, ionic bonds can produce some of the strongest bonds on the planet i.e. steel. Since covalent bonds are commonly found in carbon-based organic molecules, such as carbohydrates, our DNA and proteins are the bonds we discuss most, and covalent bonds are also found in inorganic molecules like H_2O, CO_2, and O_2, Biologists often think of covalent bonds begin the strongest. In addition, One, two, or three pairs of electrons may be shared, making single, double, and triple bonds, respectively. The more covalent bonds between two atoms, the stronger their connection. Thus, triple bonds are the strongest in biologic systems.

The strength of different levels of covalent bonding is one of the main reasons living organisms have a difficult time in acquiring nitrogen for use in constructing their molecules, even though molecular nitrogen, N_2, is the most abundant gas in the atmosphere. Molecular nitrogen consists of two nitrogen atoms triple bonded to each other and, as with all molecules, the sharing of these three pairs of electrons between the two nitrogen atoms allows for the filling of their outer electron shells, making the molecule more stable than the individual nitrogen atoms. This strong triple bond makes it difficult for living systems to break apart this nitrogen in order to use it as constituents of proteins and DNA.

The formation of water molecules provides an example of covalent bonding. The hydrogen and oxygen atoms that combine to form water molecules are bound together by covalent bonds, as shown in **Figure 10.7**. The electron from the hydrogen splits its time between the incomplete outer shell of the hydrogen atoms and the incomplete outer shell of the oxygen atoms. To completely fill the outer shell of oxygen, which has six electrons in its outer shell but which would be more stable with eight, two electrons (one from each hydrogen atom) are needed: hence the well-known formula H_2O. The electrons are shared between the two elements to fill the outer shell of each, making both elements more stable.

Polar Covalent Bonds

There are two types of covalent bonds: polar and nonpolar. In a **polar covalent bond**, shown in **Figure 10.10**, the electrons are unequally shared by the atoms and are attracted more to one nucleus than the other. Because of the unequal distribution of electrons between the atoms of different elements, a slightly positive ($\delta+$) or slightly negative ($\delta-$) charge develops. This partial charge is an important property of water and accounts for many of its characteristics.

Water is a polar molecule, with the hydrogen atoms acquiring a partial positive charge and the oxygen a partial negative charge. This occurs because the nucleus of the oxygen atom is more attractive to the electrons of the hydrogen atoms than the hydrogen nucleus is to the oxygen's electrons. Thus oxygen has a higher **electronegativity** than hydrogen and the shared electrons spend more time in the vicinity of the oxygen nucleus than they do near the nucleus of the hydrogen atoms, giving the atoms of oxygen and hydrogen slightly negative and positive charges, respectively. Another way of stating this is that the probability of finding a shared electron near an oxygen nucleus is more likely than finding it near a hydrogen nucleus. Either way, the atom's relative electronegativity contributes to the development of partial charges whenever one element is significantly more electronegative than the other, and the charges generated by these polar bonds may then be used for the formation of hydrogen bonds based on the attraction of opposite partial charges. (Hydrogen bonds, which are discussed in detail below, are weak bonds between slightly positively charged hydrogen atoms to slightly negatively charged atoms in other molecules.) Since macromolecules often have atoms within them that differ in electronegativity, polar bonds are often present in organic molecules.

Nonpolar Covalent Bonds

Nonpolar covalent bonds form between two atoms of the same element or between different elements that share electrons equally. For example, molecular oxygen (O_2) is nonpolar because the electrons will be equally distributed between the two oxygen atoms.

Another example of a nonpolar covalent bond is methane (CH_4), also shown in **Figure 10.10**. Carbon has four electrons in its outermost shell and needs four more to fill it. It gets these four from four hydrogen atoms, each atom providing one, making a stable outer shell of eight electrons. Carbon and hydrogen do not have the same electronegativity but are similar; thus, nonpolar bonds form. The hydrogen atoms each need one electron for their outermost shell, which is filled when it contains two electrons. These elements share the electrons equally among the carbons and the hydrogen atoms, creating a nonpolar covalent molecule.

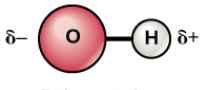
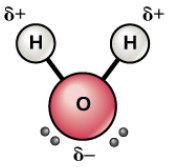
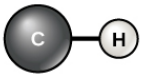
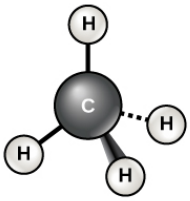
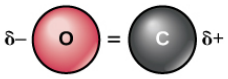
	Bond type	Molecular shape	Molecular type
Water	 Polar covalent	 Bent	Polar
Methane	 Nonpolar covalent	 Tetrahedral	Nonpolar
Carbon dioxide	 Polar covalent	 Linear	Nonpolar

Figure 10.10 Whether a molecule is polar or nonpolar depends both on bond type and molecular shape. Both water and carbon dioxide have polar covalent bonds, but carbon dioxide is linear, so the partial charges on the molecule cancel each other out.

Hydrogen Bonds and Van Der Waals Interactions

Ionic and covalent bonds between elements require energy to break. Iconic bonds are not as strong as covalent, which determines their behavior in biological systems. However, not all bonds are ionic or covalent bonds. Weaker bonds can also form between molecules. Two weak bonds that occur frequently are hydrogen bonds and van der Waals interactions. Without these two types of bonds, life as we know it would not exist. Hydrogen bonds provide many of the critical, life-sustaining properties of water and also stabilize the structures of proteins and DNA, the building block of cells.

When polar covalent bonds containing hydrogen form, the hydrogen in that bond has a slightly positive charge because hydrogen's electron is pulled more strongly toward the other element and away from the hydrogen. Because the hydrogen is slightly positive, it will be attracted to neighboring negative charges. When this happens, a weak interaction occurs between the $\delta^+$of the hydrogen from one molecule and the δ– charge on the more electronegative atoms of another molecule, usually oxygen or nitrogen, or within the same molecule. This interaction is called a **hydrogen bond**. This type of bond is common and occurs regularly between water molecules. Individual hydrogen bonds are weak and easily broken; however, they occur in very large numbers in water and in organic polymers, creating a major force in combination. Hydrogen bonds are also responsible for zipping together the DNA double helix.

Like hydrogen bonds, van der Waals interactions are weak attractions or interactions between molecules. Van der Waals attractions can occur between any two or more molecules and are dependent on slight fluctuations of the electron densities, which are not always symmetrical around an atom. For these attractions to happen, the molecules need to be very close to one another. These bonds, along with hydrogen bonds, help form the three-dimensional structure of the proteins in our cells that is necessary for their proper function.

10.2 | Water: the Molecule of Life

Introduction

> "Life can be thought of as water kept in the right temperature in the right atmosphere in the right light for a long enough period of time."

Norman J. Berrill, *You and the Universe*, pg. 45, 1958

Why do scientists spend time looking for water on other planets? Why is water so important? It is because water is essential to life as we know it. Water is one of the more abundant molecules and the one most critical to life on Earth. Approximately 70 percent of the human body is made up of water. Without it, life as we know it simply would not exist.

The **polarity** of the water molecule and its resulting **hydrogen bonding** make water a unique substance with special properties that are intimately tied to the processes of life. Life originally evolved in a watery environment, and most of an organism's cellular chemistry and metabolism occur inside the watery contents of the cell's cytoplasm. Special properties of water are its high heat capacity and heat of vaporization, its ability to dissolve polar molecules, its cohesive and adhesive properties, and its dissociation into ions that leads to the generation of pH. Understanding these characteristics of water helps to elucidate its importance in maintaining life.

Water's Polarity

One of water's important properties is that it is composed of polar molecules: the hydrogen and oxygen within water molecules (H_2O) form polar covalent bonds. While there is no net charge to a water molecule, the polarity of water creates a slightly positive charge on hydrogen and a slightly negative charge on oxygen, contributing to water's properties of attraction. Water's charges are generated because oxygen is more electronegative than hydrogen, making it more likely that a shared electron would be found near the oxygen nucleus than the hydrogen nucleus, thus generating the partial negative charge near the oxygen.

As a result of water's polarity, each water molecule attracts other water molecules because of the opposite charges between water molecules, forming hydrogen bonds. Water also attracts or is attracted to other polar molecules and ions. A polar substance that interacts readily with or dissolves in water is referred to as **hydrophilic** (hydro- = "water"; -philic = "loving"). In contrast, non-polar molecules such as oils and fats do not interact well with water, as shown in **Figure 10.11** and separate from it rather than dissolve in it, as we see in salad dressings containing oil and vinegar (an acidic water solution). These nonpolar compounds are called **hydrophobic** (hydro- = "water"; -phobic = "fearing").

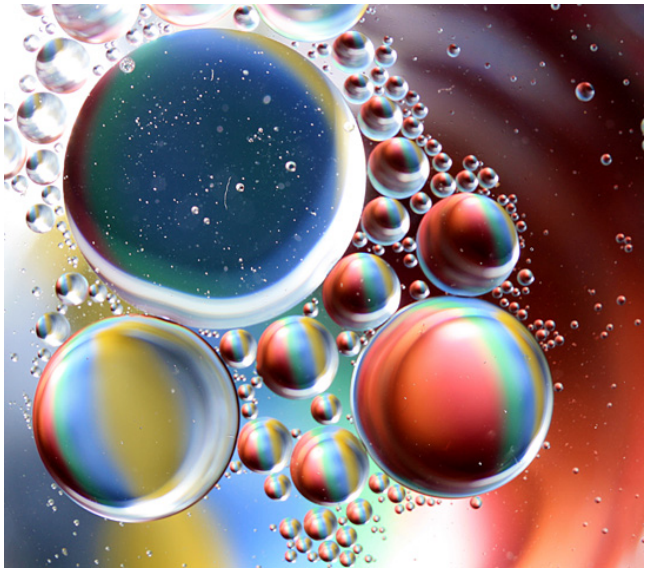


Figure 10.11 Oil and water do not mix. As this macro image of oil and water shows, oil does not dissolve in water but forms droplets instead. This is due to it being a nonpolar compound. (credit: Gautam Dogra).

Water's States: Gas, Liquid, and Solid

The formation of hydrogen bonds is an important quality of the liquid water that is crucial to life as we know it. As water molecules make hydrogen bonds with each other, water takes on some unique chemical characteristics compared to other liquids and, since living things have a high water content, understanding these chemical features is key to understanding life. In liquid water, hydrogen bonds are constantly formed and broken as the water molecules slide past each other. The breaking of these bonds is caused by the motion (kinetic energy) of the water molecules due to the heat contained in the system. When the heat is raised as water is boiled, the higher kinetic energy of the water molecules causes the hydrogen bonds to break completely and allows water molecules to escape into the air as gas (steam or water vapor). On the other hand, when the temperature of water is reduced and water freezes, the water molecules form a crystalline structure maintained by hydrogen bonding (there is not enough energy to break the hydrogen bonds) that makes ice less dense than liquid water, a

phenomenon not seen in the solidification of other liquids.

Water's lower density in its solid form is due to the way hydrogen bonds are oriented as it freezes: the water molecules are pushed farther apart compared to liquid water. With most other liquids, solidification when the temperature drops includes the lowering of kinetic energy between molecules, allowing them to pack even more tightly than in liquid form and giving the solid a greater density than the liquid.

The lower density of ice, illustrated and pictured in **Figure 10.12**, an anomaly, causes it to float at the surface of liquid water, such as in an iceberg or in the ice cubes in a glass of ice water. In lakes and ponds, ice will form on the surface of the water creating an insulating barrier that protects the animals and plant life in the pond from freezing. Without this layer of insulating ice, plants and animals living in the pond would freeze in the solid block of ice and could not survive. The detrimental effect of freezing on living organisms is caused by the expansion of ice relative to liquid water. The ice crystals that form upon freezing rupture the delicate membranes essential for the function of living cells, irreversibly damaging them. Cells can only survive freezing if the water in them is temporarily replaced by another liquid like glycerol.

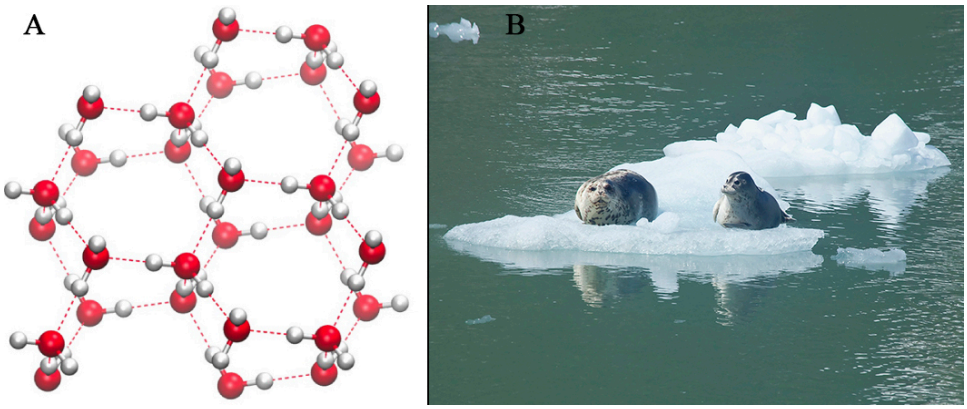


Figure 10.12 Hydrogen bonding makes ice less dense than liquid water. The (a) lattice structure of ice makes it less dense than the freely flowing molecules of liquid water, enabling it to (b) float on water. (credit a: modification of work by Jane Whitney, image created using Visual Molecular Dynamics (VMD) software[1]; credit b: courtesy David A. Rintoul)

Water's High Heat Capacity

Water's high heat capacity is a property caused by hydrogen bonding among water molecules. Water has the highest specific heat capacity of any liquids. **Specific heat** is defined as the amount of heat one gram of a substance must absorb or lose to change its temperature by one degree Celsius. For water, this amount is one calorie. It therefore takes water a long time to heat and long time to cool. In fact, the specific heat capacity of water is about five times more than that of sand. This explains why the land cools faster than the sea. Due to its high heat capacity, water is used by warm blooded animals to more evenly disperse heat in their bodies: it acts in a similar manner to a car's cooling system, transporting heat from warm places to cool places, causing the body to maintain a more even temperature.

Water's Heat of Vaporization

Water also has a high **heat of vaporization**, the amount of energy required to change one gram of a liquid substance to a gas. A considerable amount of heat energy (586 cal) is required to accomplish this change in water. This process occurs on the surface of water. As liquid water heats up, hydrogen bonding makes it difficult to separate the liquid water molecules from each other, which is required for it to enter its gaseous phase (steam). As a result, water acts as a heat sink or heat reservoir and requires much more heat to boil than does a liquid such as ethanol (grain alcohol), whose hydrogen bonding with other ethanol molecules is weaker than water's hydrogen bonding. Eventually, as water reaches its boiling point of 100° Celsius (212° Fahrenheit), the heat is able to break the hydrogen bonds between the water molecules, and the kinetic energy (motion) between the water molecules allows them to escape from the liquid as a gas. Even when below its boiling point, water's individual molecules acquire enough energy from other water molecules such that some surface water molecules can escape and vaporize: this process is known as evaporation.

The fact that hydrogen bonds need to be broken for water to evaporate means that a substantial amount of energy is used in the process. As the water evaporates, energy is taken up by the process, cooling the environment where the evaporation is taking place. In many living organisms, including in humans, the evaporation of sweat, which is 90 percent water, allows the organism to cool so that homeostasis of body temperature can be maintained.

1. W. Humphrey W., A. Dalke, and K. Schulten, "VMD—Visual Molecular Dynamics," *Journal of Molecular Graphics* 14 (1996): 33-38.

Water's Solvent Properties

Since water is a polar molecule with slightly positive and slightly negative charges, ions and polar molecules can readily dissolve in it. Therefore, water is referred to as a **solvent**, a substance capable of dissolving other polar molecules and ionic compounds. The charges associated with these molecules will form hydrogen bonds with water, surrounding the particle with water molecules. This is referred to as a sphere of hydration, or a hydration shell, as illustrated in **Figure 10.13** and serves to keep the particles separated or dispersed in the water.

When ionic compounds are added to water, the individual ions react with the polar regions of the water molecules and their ionic bonds are disrupted in the process of dissociation. Dissociation occurs when atoms or groups of atoms break off from molecules and form ions. Consider table salt (NaCl, or sodium chloride): when NaCl crystals are added to water, the molecules of NaCl dissociate into Na^+ and Cl^- ions, and spheres of hydration form around the ions, illustrated in **Figure 10.13**. The positively charged sodium ion is surrounded by the partially negative charge of the water molecule's oxygen. The negatively charged chloride ion is surrounded by the partially positive charge of the hydrogen on the water molecule.

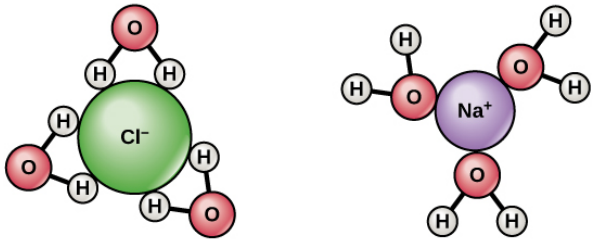


Figure 10.13 When table salt (NaCl) is mixed in water, spheres of hydration are formed around the ions.

Water's Cohesive and Adhesive Properties

Have you ever filled a glass of water to the very top and then slowly added a few more drops? Before it overflows, the water forms a dome-like shape above the rim of the glass. This water can stay above the glass because of the property of **cohesion**. In cohesion, water molecules are attracted to each other (because of hydrogen bonding), keeping the molecules together at the liquid-gas (water-air) interface, although there is no more room in the glass.

Cohesion allows for the development of surface tension, the capacity of a substance to withstand being ruptured when placed under tension or stress. This is also why water forms droplets when placed on a dry surface rather than being flattened out by gravity. When a small scrap of paper is placed onto the droplet of water, the paper floats on top of the water droplet even though paper is denser (heavier) than the water. Cohesion and surface tension keep the hydrogen bonds of water molecules intact and support the item floating on the top. It's even possible to "float" a needle on top of a glass of water if it is placed gently without breaking the surface tension, as shown in **Figure 10.14**.

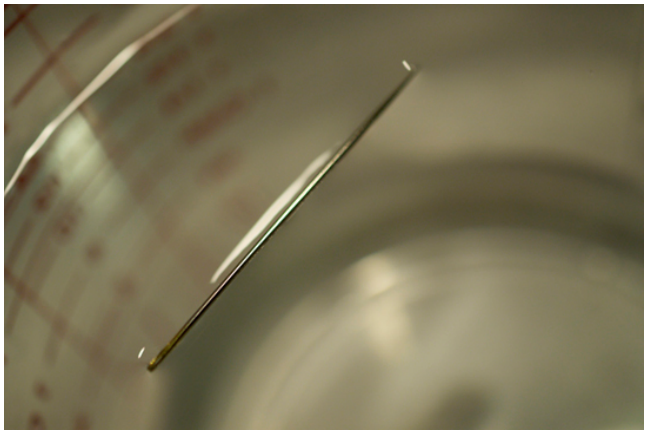


Figure 10.14 The weight of the needle is pulling the surface downward; at the same time, the surface tension is pulling it up, suspending it on the surface of the water and keeping it from sinking. Notice the indentation in the water around the needle. (credit: Cory Zanker)

These cohesive forces are related to water's property of **adhesion**, or the attraction between water molecules and other molecules. This attraction is sometimes stronger than water's cohesive forces, especially when the water is exposed to charged surfaces such as those found on the inside of thin glass tubes known as capillary tubes. Adhesion is observed when

water “climbs” up the tube placed in a glass of water: notice that the water appears to be higher on the sides of the tube than in the middle. This is because the water molecules are attracted to the charged glass walls of the capillary more than they are to each other and therefore adhere to it. This type of adhesion is called capillary action, and is illustrated in **Figure 10.15**.

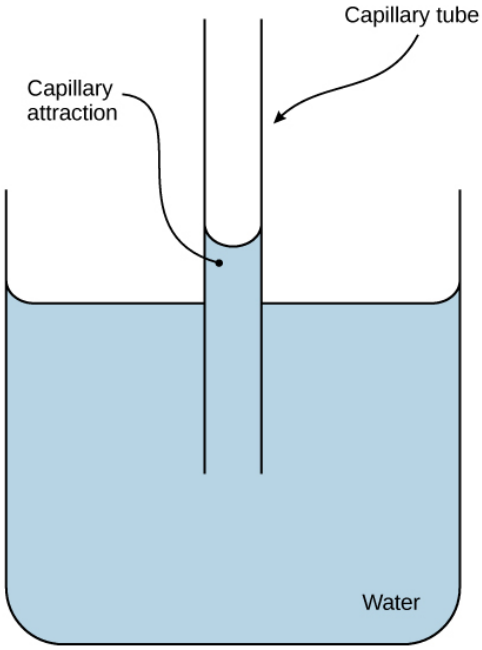


Figure 10.15 Capillary action in a glass tube is caused by the adhesive forces exerted by the internal surface of the glass exceeding the cohesive forces between the water molecules themselves. (credit: modification of work by Pearson-Scott Foresman, donated to the Wikimedia Foundation)

Why are cohesive and adhesive forces important for life? Cohesive and adhesive forces are important for the transport of water from the roots to the leaves in plants. These forces create a “pull” on the water column. This pull results from the tendency of water molecules being evaporated on the surface of the plant to stay connected to water molecules below them, and so they are pulled along. Plants use this natural phenomenon to help transport water from their roots to their leaves. Without these properties of water, plants would be unable to receive the water and the dissolved minerals they require. In another example, insects such as the water strider, shown in **Figure 10.16**, use the surface tension of water to stay afloat on the surface layer of water and even mate there.

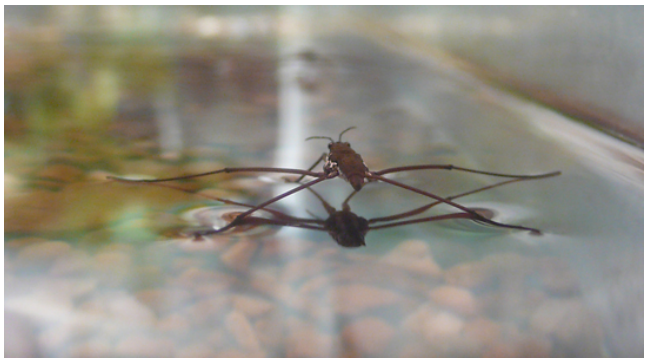


Figure 10.16 Water’s cohesive and adhesive properties allow this water strider (*Gerris* sp.) to stay afloat. (credit: Tim Vickers)

pH, Acids, and Bases

The pH of a solution indicates its acidity or alkalinity.

$$H_2O(I) \leftrightarrow H^+(aq) + OH^-(aq)$$

Litmus or pH paper, filter paper that has been treated with a natural water-soluble dye so it can be used as a pH indicator,

to test how much acid (acidity) or base (alkalinity) exists in a solution. You might have even used some to test whether the water in a swimming pool is properly treated. In both cases, the pH test measures the concentration of hydrogen ions in a given solution.

Hydrogen ions are spontaneously generated in pure water by the dissociation (ionization) of a small percentage of water molecules into equal numbers of hydrogen (H^+) ions and hydroxide (OH^-) ions. While the hydroxide ions are kept in solution by their hydrogen bonding with other water molecules, the hydrogen ions, consisting of naked protons, are immediately attracted to un-ionized water molecules, forming hydronium ions (H_30^+). Still, by convention, scientists refer to hydrogen ions and their concentration as if they were free in this state in liquid water.

The concentration of hydrogen ions dissociating from pure water is 1×10^{-7} moles H^+ ions per liter of water. Moles (mol) are a way to express the amount of a substance (which can be atoms, molecules, ions, etc), with one mole being equal to 6.02×10^{23} particles of the substance. Therefore, 1 mole of water is equal to 6.02×10^{23} water molecules. The pH is calculated as the negative of the base 10 logarithm of this concentration (see below). The log10 of 1×10^{-7} is -7.0, and the negative of this number (indicated by the “p” of “pH”) yields a pH of 7.0, which is also known as neutral pH. The pH inside of human cells and blood are examples of two areas of the body where near-neutral pH is maintained.

pH = -log10[Hydrogen Ion]

Non-neutral pH readings result from dissolving acids or bases in water. Using the negative logarithm to generate positive integers, high concentrations of hydrogen ions yield a low pH number, whereas low levels of hydrogen ions result in a high pH. An **acid** is a substance that increases the concentration of hydrogen ions (H^+) in a solution, usually by having one of its hydrogen atoms dissociate. A **base** provides either hydroxide ions (OH^-) or other negatively charged ions that combine with hydrogen ions, reducing their concentration in the solution and thereby raising the pH. In cases where the base releases hydroxide ions, these ions bind to free hydrogen ions, generating new water molecules.

The stronger the acid, the more readily it donates H^+. For example, hydrochloric acid (HCl) completely dissociates into hydrogen and chloride ions and is highly acidic, whereas the acids in tomato juice or vinegar do not completely dissociate and are considered weak acids. Conversely, strong bases are those substances that readily donate OH^- or take up hydrogen ions. Sodium hydroxide (NaOH) and many household cleaners are highly alkaline and give up OH^- rapidly when placed in water, thereby raising the pH. An example of a weak basic solution is seawater, which has a pH near 8.0, close enough to neutral pH that marine organisms adapted to this saline environment are able to thrive in it.

The **pH scale** is, as previously mentioned, an inverse logarithm and ranges from 0 to 14 (**Figure 10.17**). Anything below 7.0 (ranging from 0.0 to 6.9) is acidic, and anything above 7.0 (from 7.1 to 14.0) is alkaline. Extremes in pH in either direction from 7.0 are usually considered inhospitable to life. The pH inside cells (6.8) and the pH in the blood (7.4) are both very close to neutral. However, the environment in the stomach is highly acidic, with a pH of 1 to 2. So how do the cells of the stomach survive in such an acidic environment? How do they homeostatically maintain the near neutral pH inside them? The answer is that they cannot do it and are constantly dying. New stomach cells are constantly produced to replace dead ones, which are digested by the stomach acids. It is estimated that the lining of the human stomach is completely replaced every seven to ten days.

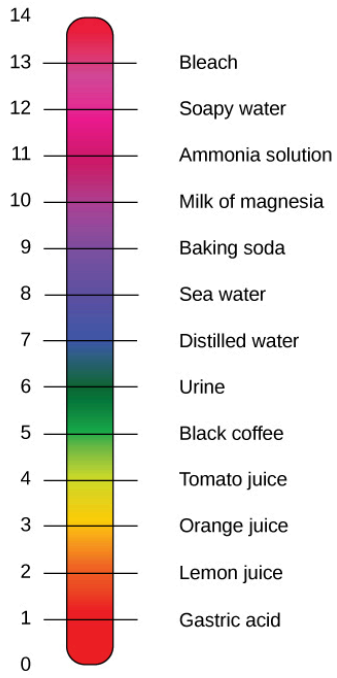


Figure 10.17 The pH scale measures the concentration of hydrogen ions (H^+) in a solution. (credit: modification of work by Edward Stevens)

10.3 | Introduction to Biological Molecules

Introduction

> "The most fundamental difference between compounds of low molecular weight and macromolecular compounds resides in the fact that the latter may exhibit properties that cannot be deduced from a close examination of the low molecular weight materials. Not very different structures can be obtained from a few building blocks, but if 10,000 or 100,000 bloocks are at hand, the most varied structures become possible..."

Hermann Staudinger, quoted in R. Oesper, *The Human Side of Scientists*, pg. 75, 1975

Biological macromolecules are large molecules, necessary for life (Table 1), and these large molecules are made from smaller organic molecules. As noted above, there is an almost infinite variety of possible structures for macromolecules, since both the composition and the bonds lining the components can vary tremendously. There are four major classes of biological macromolecules (carbohydrates, lipids, proteins, and nucleic acids). As we explore these molecules, think about how each class makes up important cell components and perform a wide array of functions. Since all these biological macromolecules contain carbon, we call them are organic molecules. In addition, the ratio of carbon, hydrogen, oxygen, nitrogen, and additional minor elements determines the class of biological molecules.

Type of Molecule	Function	Location in Cell

Table 10.3 A summary of the Biological Molecules

Carbohydrates		
Simple Sugars	Provide Quick Energy	Cytoplasm
Complex Carbohydrates (cellulose, chitin, starch, glycogen)	Support Cells (cellulose, chitin); Store energy (starch, glycogen)	Cell Walls (cellulose, chitin); cytoplasm (starch, glycogen)
Lipids		
Triglycerides (fats and oils)	Store Energy	Cytoplasm
Phospholipids	Major component of biological membranes	Plasma membranes
Sterols and Steroids	Stabilize animal plasma membranes; sex hormones	animal plasma membranes
Waxes	Waterproofing	Cell Walls (plants),Excreted (animals)
Proteins	Movement, Immunity, Energy Source, Enzymes, Structural Support, Communication, Hormones,	In all parts of the cell
Nucleic Acids (DNA and RNA)	Store and use genetic information	DNA (Nucleus) RNA (Cytoplasm, and Rough Endoplasmic Reticulum

Table 10.3 A summary of the Biological Molecules

10.4 | Chemical Reactions of Biological Macromolecules

Condensation Reaction

Most macromolecules are made from single subunits, or building blocks, called **monomers**. The monomers combine with each other using covalent bonds to form larger molecules known as **polymers**. In doing so, monomers release water molecules as byproducts. This type of reaction is known as **condensation reaction**.

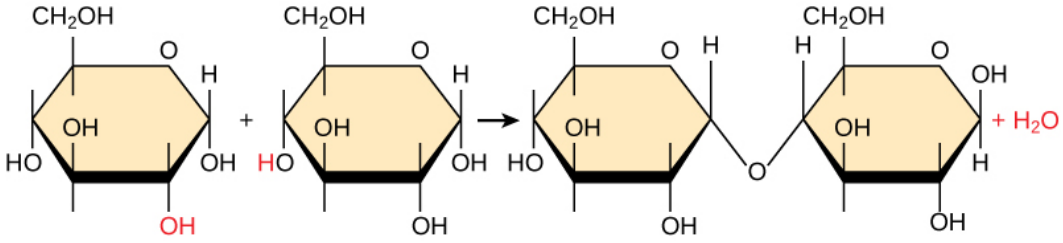


Figure 10.18 In the condensation reaction depicted above, two molecules of glucose are linked together to form the dissacharide maltose. In the process, a water molecule is formed.

In a condensation reaction (**Figure 10.18**), the hydrogen of one monomer combines with the hydroxyl group of another monomer, releasing a molecule of water. At the same time, the monomers share electrons and form covalent bonds. As additional monomers join, this chain of repeating monomers forms a polymer. Different types of monomers can combine in many configurations, giving rise to a diverse group of macromolecules. Even one kind of monomer can combine in a variety of ways to form several different polymers: for example, glucose monomers are the constituents of starch, glycogen, and cellulose.

Hydrolysis

Polymers are broken down into monomers in a process known as **hydrolysis**, which means “to split water,” a reaction in which a water molecule is used during the breakdown (Figure 10.19). During these reactions, the polymer is broken into two components: one part gains a hydrogen atom (H+) and the other gains a hydroxyl molecule (OH–) from a split water molecule.

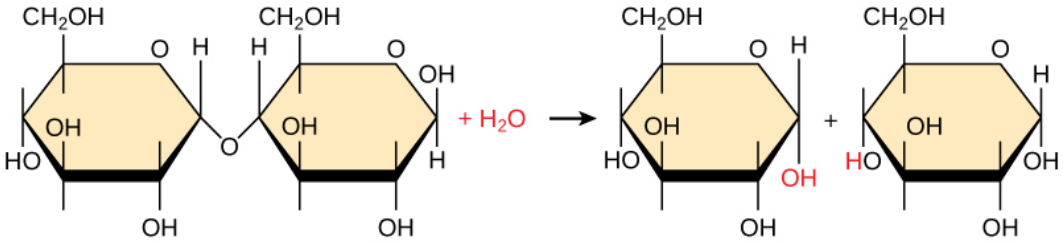


Figure 10.19 In the hydrolysis reaction shown here, the disaccharide maltose is broken down to form two glucose monomers. Note that this reaction is the reverse of the condensation reaction shown in Figure 10.18.

Condensation and **hydrolysis reactions** are catalyzed, or “sped up,” by specific enzymes; condensation reactions involve the formation of new bonds, requiring energy, while hydrolysis reactions break bonds and release energy. These reactions are similar for most macromolecules, but each monomer and polymer reaction is specific for its class. For example, in our bodies, food is hydrolyzed, or broken down, into smaller molecules by catalytic enzymes in the digestive system. This allows for easy absorption of nutrients by cells in the intestine. Each macromolecule is broken down by a specific enzyme. For instance, carbohydrates are broken down by amylase, sucrase, lactase, or maltase. Proteins are broken down by the enzymes pepsin and peptidase, and by hydrochloric acid. Lipids are broken down by lipases. Breakdown of these macromolecules provides energy for cellular activities.

10.5 | Carbohydrates

Introduction

> “I called it ignose, not knowing which carbohydrate it was. This name was turned down by my editor. 'God-nose' was not more successful, so in the end 'hexuronic acid' was agreed upon. To-day the substance is called 'ascorbic acid' and I will use this name.”

Albert Szent-Gyorgyi, *Studies on Biological Oxidation and Some of its Catalysts (C4 Dicarboxylic Acids, Vitamin C and P Etc.)*, pg. 73, 1937

Most people(except perhaps Szent-Gyorgyi's editor) are familiar with carbohydrates, one type of macromolecule, especially when it comes to what we eat. To lose weight, some individuals adhere to “low-carb” diets. Athletes, in contrast, often “carb-load” before important competitions to ensure that they have enough energy to compete at a high level. Carbohydrates are, in fact, an essential part of our diet; grains, fruits, and vegetables are all natural sources of carbohydrates. Carbohydrates provide energy to the body, particularly through glucose, a simple sugar that is a component of **starch** and an ingredient in many staple foods. Carbohydrates also have other important functions in humans, animals, and plants. And, as you can infer from the epigraph above, the suffix "-ose" is a good clue that a word describes a carboydrate of some sort.

Molecular Structures

Carbohydrates can be represented by the stoichiometric formula $(CH_2O)_n$, where n is the number of carbons in the molecule. In other words, the ratio of carbon to hydrogen to oxygen is 1:2:1 in carbohydrate molecules. This formula also explains the origin of the term “carbohydrate”: the components are carbon (“carbo”) and the components of water (hence, “hydrate”). Carbohydrates are classified into three subtypes: monosaccharides, disaccharides, and polysaccharides.

Monosaccharides

Monosaccharides (mono- = “one”; sacchar- = “sweet”) are simple sugars, the most common of which is glucose. In monosaccharides, the number of carbons usually ranges from three to seven. Most monosaccharide names end with the suffix -ose. Depending on the number of carbons in the sugar, they also may be known as trioses (three carbons), pentoses (five carbons), and or hexoses (six carbons). See Figure 10.20 for an illustration of the monosaccharides.

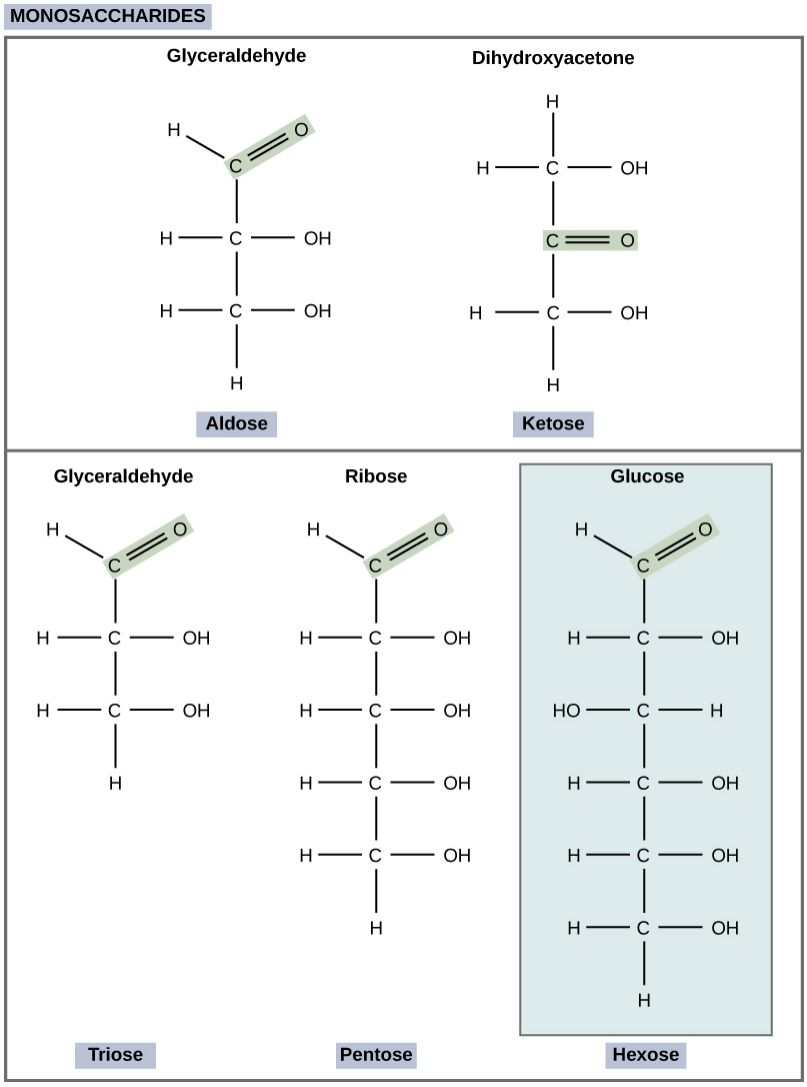


Figure 10.20 Monosaccharides are classified based on the position of their carbonyl group and the number of carbons in the backbone. Aldoses have a carbonyl group (indicated in green) at the end of the carbon chain, and ketoses have a carbonyl group in the middle of the carbon chain. Trioses, pentoses, and hexoses have three, five, and six carbon backbones, respectively.

The chemical formula for glucose is $C_6H_{12}O_6$. In humans, glucose is an important source of energy. During cellular respiration, energy is released from glucose, and that energy is used to help make adenosine triphosphate (ATP). Plants synthesize glucose using carbon dioxide and water, and glucose in turn is used for energy requirements for the plant. Excess glucose is often stored as starch that is catabolized (the breakdown of larger molecules by cells) by humans and other animals that feed on plants.

Galactose (part of lactose, or milk sugar) and fructose (found in sucrose, in fruit) are other common monosaccharides. Although glucose, galactose, and fructose all have the same chemical formula ($C_6H_{12}O_6$), they differ structurally and chemically (and are known as isomers) because of the different arrangement of functional groups around the asymmetric carbon; all of these monosaccharides have more than one asymmetric carbon (Figure 10.21).

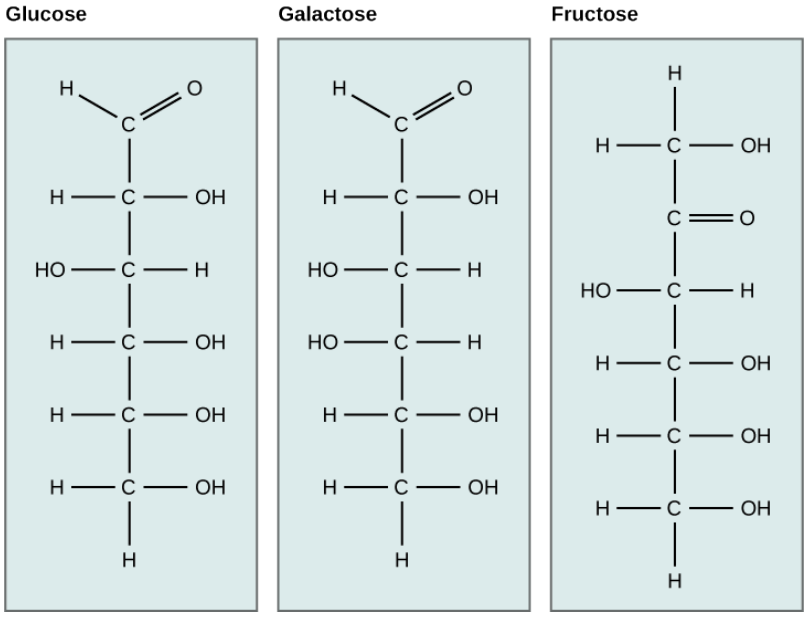


Figure 10.21 Glucose, galactose, and fructose are all hexoses. They are structural isomers, meaning they have the same chemical formula ($C_6H_{12}O_6$) but a different arrangement of atoms.

Monosaccharides can exist as a linear chain or as ring-shaped molecules; in aqueous solutions they are usually found in ring forms (**Figure 10.22**). Glucose in a ring form can have two different arrangements of the hydroxyl group (OH) around the anomeric carbon (carbon 1 that becomes asymmetric in the process of ring formation). If the hydroxyl group is below carbon number 1 in the sugar, it is said to be in the alpha (α) position, and if it is above the plane, it is said to be in the beta (β) position.

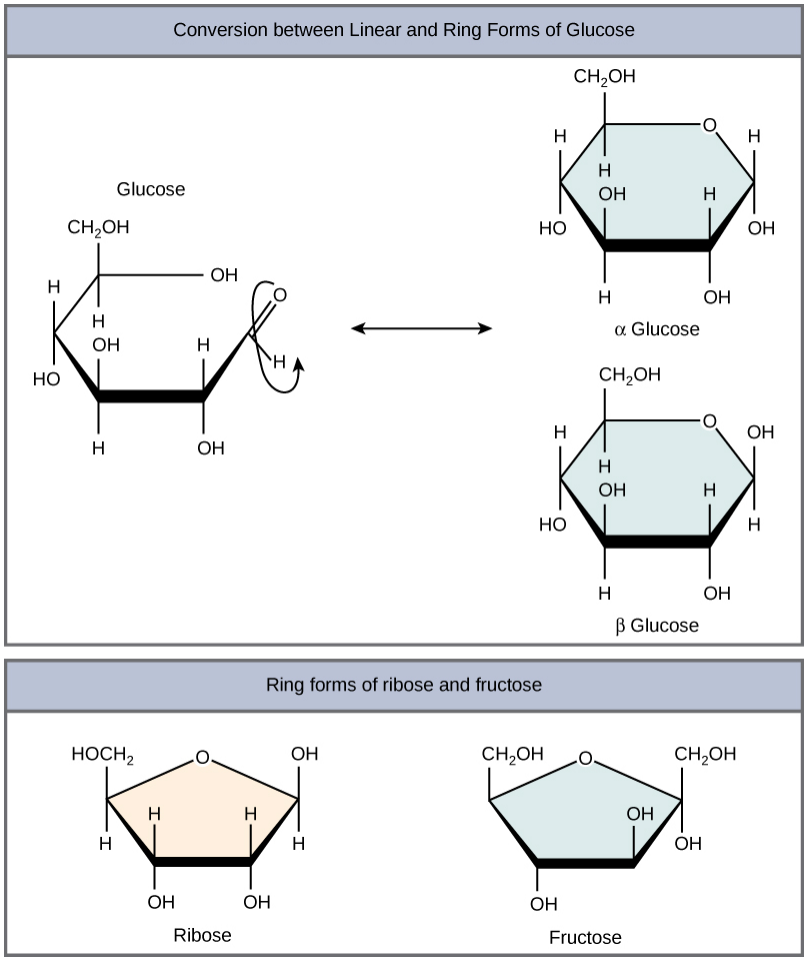


Figure 10.22 Five and six carbon monosaccharides exist in equilibrium between linear and ring forms. When the ring forms, the side chain it closes on is locked into an *α* or *β* position. Fructose and ribose also form rings, although they form five-membered rings as opposed to the six-membered ring of glucose.

Disaccharides

Disaccharides (di- = “two”) form when two monosaccharides undergo a condensation reaction (also known as a dehydration synthesis). During this process, the hydroxyl group of one monosaccharide combines with the hydrogen of another monosaccharide, releasing a molecule of water and forming a covalent bond. A covalent bond formed between a carbohydrate molecule and another molecule (in this case, between two monosaccharides) is known as a **glycosidic bond** (**Figure 10.23**). Glycosidic bonds (also called glycosidic linkages) can be of the alpha or the beta type.

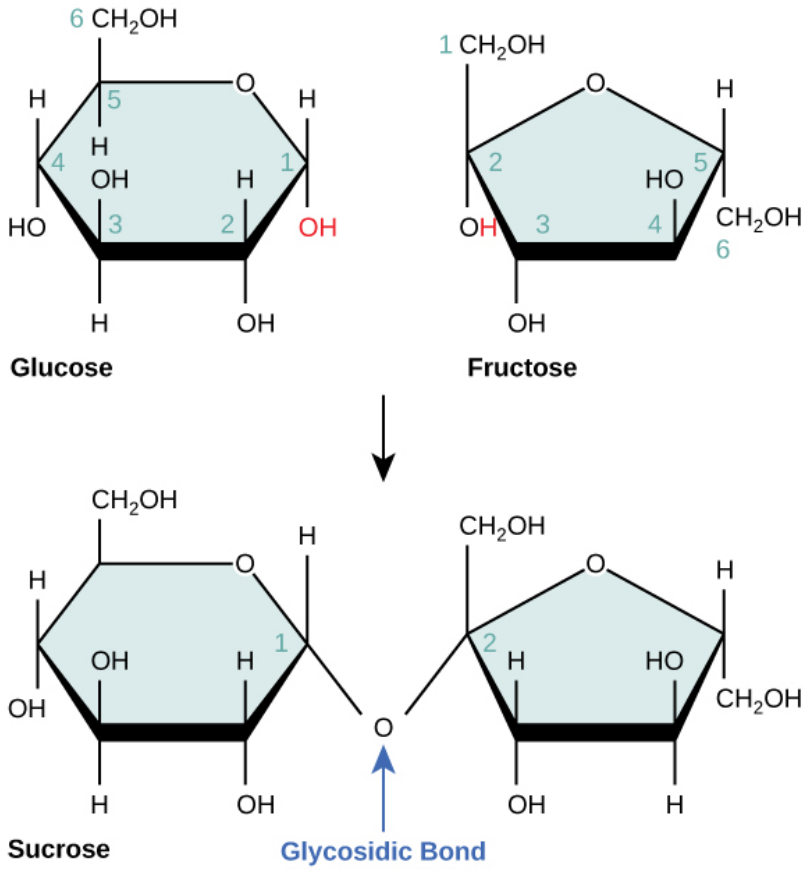


Figure 10.23 Sucrose is formed when a monomer of glucose and a monomer of fructose are joined in a condensation reaction to form a glycosidic bond. In the process, a water molecule is lost. By convention, the carbon atoms in a monosaccharide are numbered from the terminal carbon closest to the carbonyl group. In sucrose, a glycosidic linkage is formed between carbon 1 in glucose and carbon 2 in fructose.

Common disaccharides include lactose, maltose, and sucrose (**Figure 10.24**). Lactose is a disaccharide consisting of the monomers glucose and galactose. It is found naturally in milk. Maltose, or malt sugar, is a disaccharide formed by a condensation reaction between two glucose molecules. The most common disaccharide is sucrose, or table sugar, which is composed of the monomers glucose and fructose.

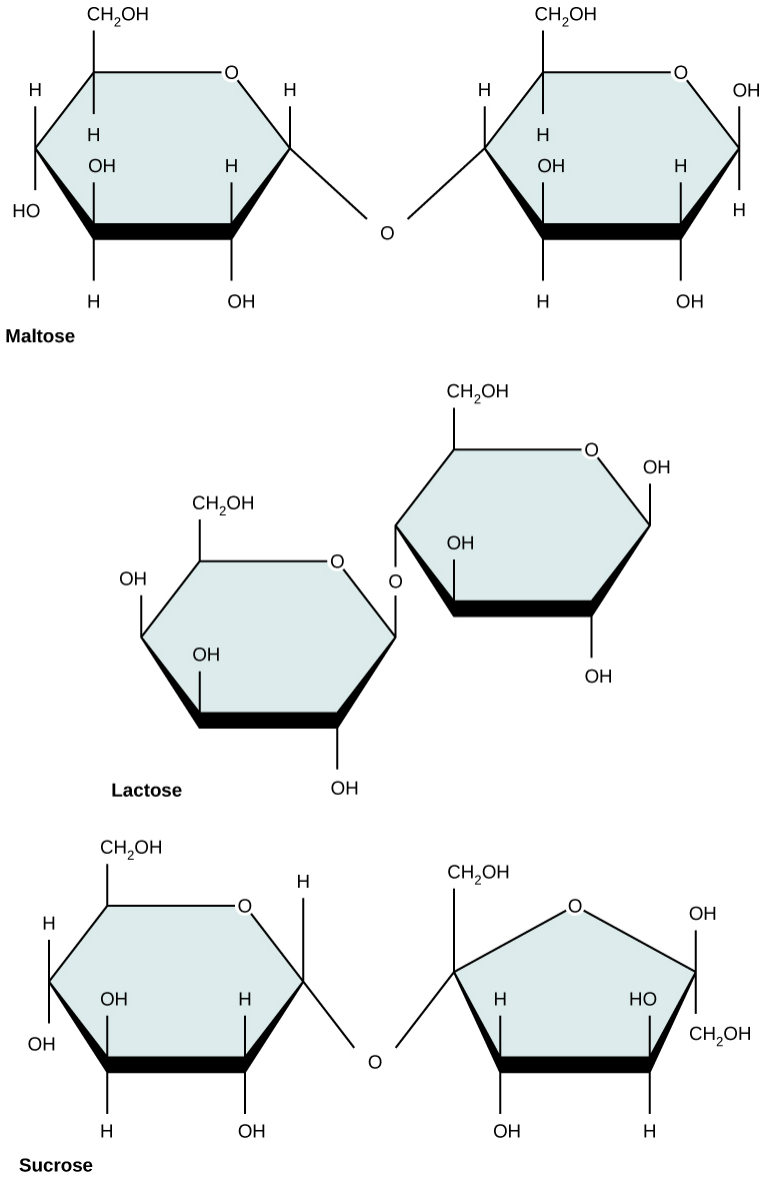


Figure 10.24 Common disaccharides include maltose (grain sugar), lactose (milk sugar), and sucrose (table sugar).

Polysaccharides

A long chain of monosaccharides linked by glycosidic bonds is known as a **polysaccharide** (poly- = "many"). The chain may be branched or unbranched, and it may contain different types of monosaccharides. The molecular weight may be 100,000 daltons or more depending on the number of monomers joined. Starch, glycogen, cellulose, and chitin are primary examples of polysaccharides.

Starch is the stored form of sugars in plants and is made up of a mixture of amylose and amylopectin (both polymers of glucose). Plants are able to synthesize glucose, and the excess glucose, beyond the plant's immediate energy needs, is stored as starch in different plant parts, including roots and seeds. The starch in the seeds provides food for the embryo as it germinates and can also act as a source of food for humans and animals. The starch that is consumed by humans is broken down by enzymes, such as salivary amylases, into smaller molecules, such as maltose and glucose. The cells can then absorb the glucose.

Starch is made up of glucose monomers that are joined by α 1-4 or α 1-6 glycosidic bonds. The numbers 1-4 and 1-6 refer to the carbon number of the two residues that have joined to form the bond. As illustrated in **Figure 10.25**, amylose is starch formed by unbranched chains of glucose monomers (only α 1-4 linkages), whereas amylopectin is a branched polysaccharide (α 1-6 linkages at the branch points).

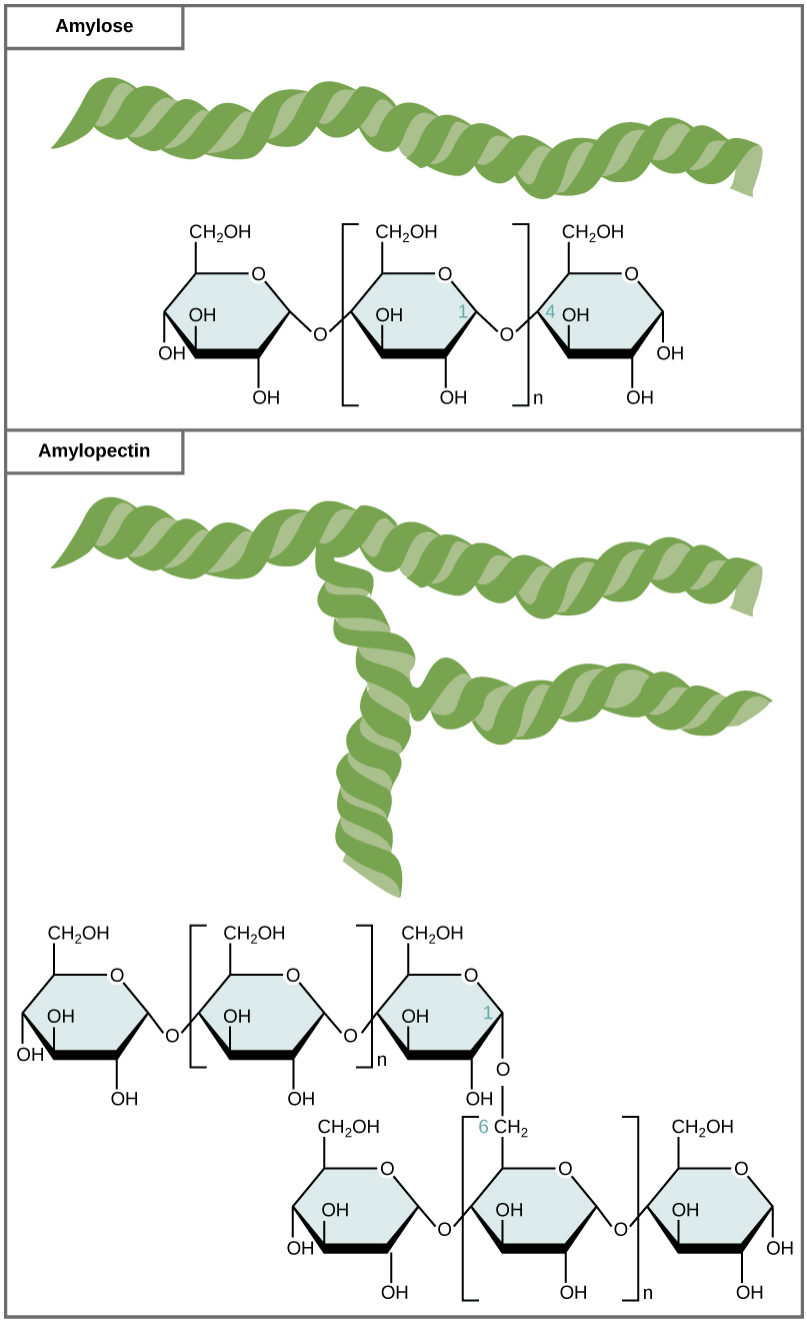


Figure 10.25 Amylose and amylopectin are two different forms of starch. Amylose is composed of unbranched chains of glucose monomers connected by *α* 1,4 glycosidic linkages. Amylopectin is composed of branched chains of glucose monomers connected by *α* 1,4 and *α* 1,6 glycosidic linkages. Because of the way the subunits are joined, the glucose chains have a helical structure. Glycogen (not shown) is similar in structure to amylopectin but more highly branched.

Glycogen is the storage form of glucose in humans and other vertebrates and is made up of monomers of glucose. Glycogen is the animal equivalent of starch and is a highly branched molecule usually stored in liver and muscle cells. Whenever blood glucose levels decrease, glycogen is broken down to release glucose in a process known as glycogenolysis.

Cellulose is the most abundant natural biopolymer. The cell wall of plants is mostly made of cellulose; this provides structural support to the cell. Wood and paper are mostly cellulosic in nature. Cellulose is made up of glucose monomers that are linked by *β* 1-4 glycosidic bonds (**Figure 10.26**).

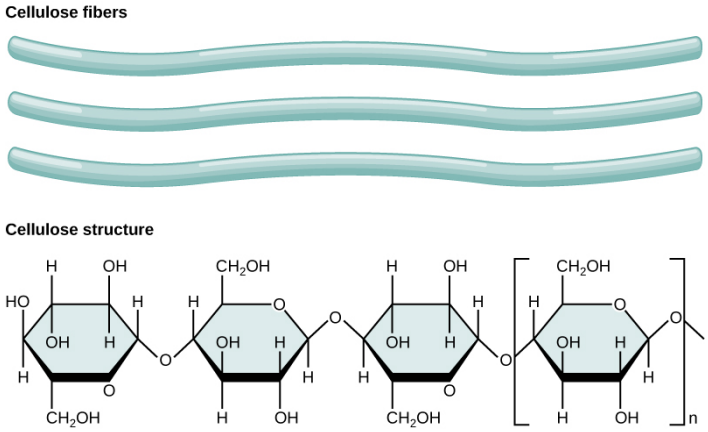


Figure 10.26 In cellulose, glucose monomers are linked in unbranched chains by β 1-4 glycosidic linkages. Because of the way the glucose subunits are joined, every glucose monomer is flipped relative to the next one resulting in a linear, fibrous structure.

As shown in **Figure 10.26**, every other glucose monomer in cellulose is flipped over, and the monomers are packed tightly as extended long chains. This gives cellulose its rigidity and high tensile strength—which is so important to plant cells. While the β 1-4 linkage cannot be broken down by human digestive enzymes, herbivores such as cows, koalas, buffalos, and horses are able, with the help of the specialized flora in their stomach, to digest plant material that is rich in cellulose and use it as a food source. In these animals, certain species of bacteria and protists reside in the rumen (part of the digestive system of herbivores) and secrete the enzyme cellulase. The appendix of grazing animals also contains bacteria that digest cellulose, giving it an important role in the digestive systems of ruminants. Cellulases can break down cellulose into glucose monomers that can be used as an energy source by the animal. Termites are also able to break down cellulose because of the presence of other organisms in their bodies that secrete cellulases.

Carbohydrates serve various functions in different animals. Arthropods (insects, crustaceans, and others) have an outer skeleton, called the exoskeleton, which protects their internal body parts (as seen in the bee in **Figure 10.27**). This exoskeleton is made of the biological macromolecule **chitin**, which is a polysaccharide-containing nitrogen. It is made of repeating units of N-acetyl-β-d-glucosamine, a modified sugar. Chitin is also a major component of fungal cell walls; fungi are neither animals nor plants and form a kingdom of their own in the domain Eukarya.

Figure 10.27 Insects have a hard outer exoskeleton made of chitin, a type of polysaccharide. (credit: Louise Docker)

Benefits of Carbohydrates

Are carbohydrates good for you? People who wish to lose weight are often told that carbohydrates are bad for them and should be avoided. Some diets completely forbid carbohydrate consumption, claiming that a low-carbohydrate diet helps people to lose weight faster. However, carbohydrates have been an important part of the human diet for thousands of years; artifacts from ancient civilizations show the presence of wheat, rice, and corn in our ancestors' storage areas.

Carbohydrates should be supplemented with proteins, vitamins, and fats to be parts of a well-balanced diet. Calorie-wise, a gram of carbohydrate provides 4.3 Kcal. For comparison, fats provide 9 Kcal/g, a less desirable ratio. Carbohydrates contain soluble and insoluble elements; the insoluble part is known as fiber, which is mostly cellulose. Fiber has many uses; it promotes regular bowel movement by adding bulk, and it regulates the rate of consumption of blood glucose. Fiber also helps to remove excess cholesterol from the body: fiber binds to the cholesterol in the small intestine, then attaches to the cholesterol and prevents the cholesterol particles from entering the bloodstream, and then cholesterol exits the body via the feces. Fiber-rich diets also have a protective role in reducing the occurrence of colon cancer. In addition, a meal containing whole grains and vegetables gives a feeling of fullness. As an immediate source of energy, glucose is broken down during the process of cellular respiration, which produces ATP, the energy currency of the cell. Without the consumption of carbohydrates, the availability of "instant energy" would be reduced. Eliminating carbohydrates from the diet is not the best way to lose weight. A low-calorie diet that is rich in whole grains, fruits, vegetables, and lean meat, together with plenty of exercise and plenty of water, is the more sensible way to lose weight.

10.6 | Lipids

Introduction

> "A story about the Jack Spratts of medicine [was] told recently by Dr. Charles H. Best, co-discoverer of insulin. He had been invited to a conference of heart specialists in North America. On the eve of the meeting, out of respect for the fat-clogs-the-arteries theory, the delegates sat down to a special banquet served without fats. It was unpalatable but they all ate it as a duty. Next morning Best looked round the breakfast room and saw these same specialists — all in the 40-60 year old, coronary age group — happily tucking into eggs, bacon, buttered toast and coffee with cream."

Richard Mackarness, *Objections To High-Fat Diets', Eat Fat And Grow Slim*, chapter 3, 1958

Lipids include a diverse group of compounds, including those tasty fats in your diet, that are largely nonpolar in nature. This is because they are hydrocarbons that include mostly nonpolar carbon–carbon or carbon–hydrogen bonds. Non-polar molecules are hydrophobic ("water fearing"), or insoluble in water. Lipids perform many different functions in a cell. Cells store energy for long-term use in the form of fats. Lipids also provide insulation from the environment for plants and animals (**Figure 10.28**). For example, they help keep aquatic birds and mammals dry when forming a protective layer over fur or feathers because of their water-repellant hydrophobic nature. Lipids are also the building blocks of many hormones and are an important constituent of all cellular membranes. Lipids include fats, oils, waxes, phospholipids, and steroids.

Figure 10.28 Hydrophobic lipids in the fur of aquatic mammals, such as this river otter, protect them from the elements. (credit: Ken Bosma)

Fats and Oils

A fat molecule consists of two main components—glycerol and fatty acids. Glycerol is an organic compound (alcohol) with three carbons, five hydrogens, and three hydroxyl (OH) groups. Fatty acids have a long chain of hydrocarbons to which a carboxyl group is attached, hence the name "fatty acid." The number of carbons in the fatty acid may range from 4 to 36; most common are those containing 12–18 carbons. In a fat molecule, the fatty acids are attached to each of the three carbons of the glycerol molecule with an ester bond through an oxygen atom (**Figure 10.29**).

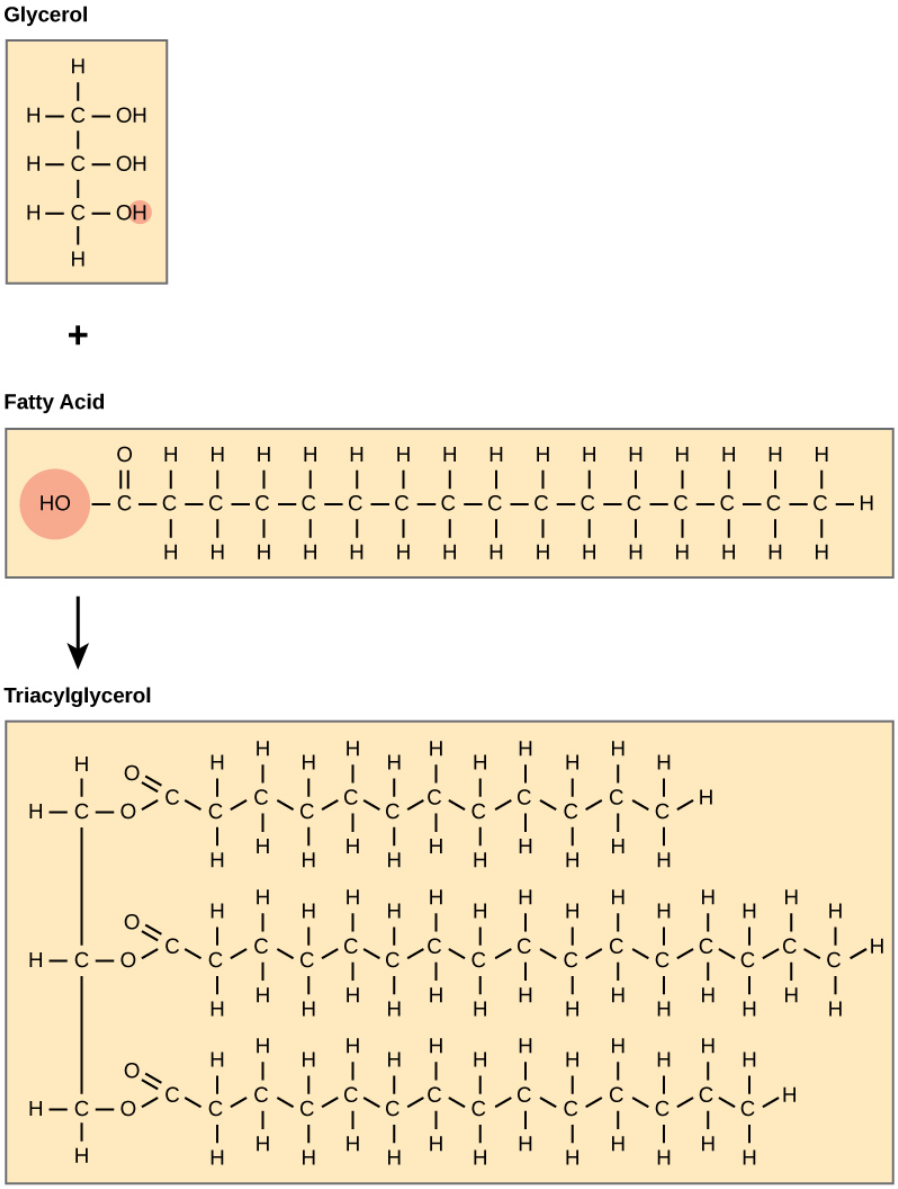


Figure 10.29 Triacylglycerol is formed by the joining of three fatty acids to a glycerol backbone in a condensation reaction. Three molecules of water are released in the process.

During this ester bond formation, three water molecules are released. The three fatty acids in the triacylglycerol may be similar or dissimilar. Fats are also called **triacylglycerols** or **triglycerides** because of their chemical structure. Some fatty acids have common names that specify their origin. For example, palmitic acid, a **saturated fatty acid**, is derived from the palm tree. Arachidic acid is derived from *Arachis hypogea*, the scientific name for groundnuts or peanuts.

Fatty acids may be saturated or unsaturated. In a fatty acid chain, if there are only single bonds between neighboring carbons in the hydrocarbon chain, the fatty acid is said to be saturated. Saturated fatty acids are saturated with hydrogen; in other words, the number of hydrogen atoms attached to the carbon skeleton is maximized. Stearic acid is an example of a saturated fatty acid (**Figure 10.30**)

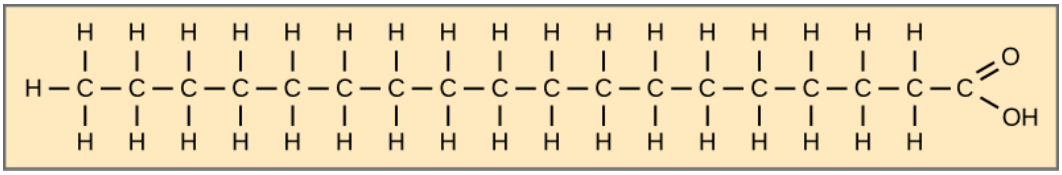


Figure 10.30 Stearic acid is a common saturated fatty acid.

When the hydrocarbon chain contains a double bond, the fatty acid is said to be **unsaturated**. Oleic acid is an example of an unsaturated fatty acid (**Figure 10.31**).

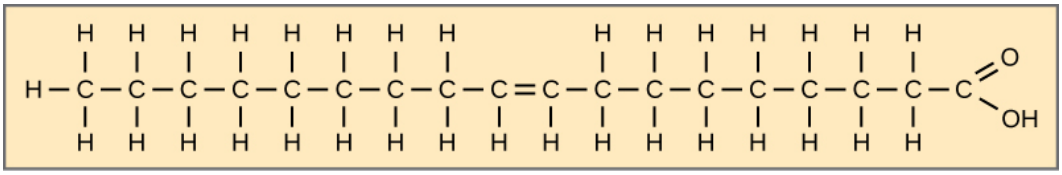


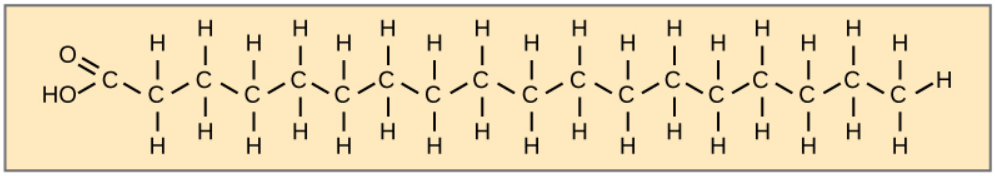
Figure 10.31 Oleic acid is a common unsaturated fatty acid.

Most unsaturated fats are liquid at room temperature and are called oils. If there is one double bond in the molecule, then it is known as a monounsaturated fat (e.g., olive oil), and if there is more than one double bond, then it is known as a polyunsaturated fat (e.g., canola oil).

When a fatty acid has no double bonds, it is known as a saturated fatty acid because no more hydrogen may be added to the carbon atoms of the chain. A fat may contain similar or different fatty acids attached to glycerol. Long straight fatty acids with single bonds tend to get packed tightly and are solid at room temperature. Animal fats with stearic acid and palmitic acid (common in meat) and the fat with butyric acid (common in butter) are examples of saturated fats. Mammals store fats in specialized cells called adipocytes, where globules of fat occupy most of the cell's volume. In plants, fat or oil is stored in many seeds and is used as a source of energy during seedling development. Unsaturated fats or oils are usually of plant origin and contain *cis* unsaturated fatty acids. *Cis* and *trans* indicate the configuration of the molecule around the double bond. If hydrogens are present in the same plane, it is referred to as a cis fat; if the hydrogen atoms are on two different planes, it is referred to as a trans fat. The *cis* double bond causes a bend or a "kink" that prevents the fatty acids from packing tightly, keeping them liquid at room temperature (**Figure 10.32**). Olive oil, corn oil, canola oil, and cod liver oil are examples of unsaturated fats. Unsaturated fats help to lower blood cholesterol levels whereas saturated fats contribute to plaque formation in the arteries.

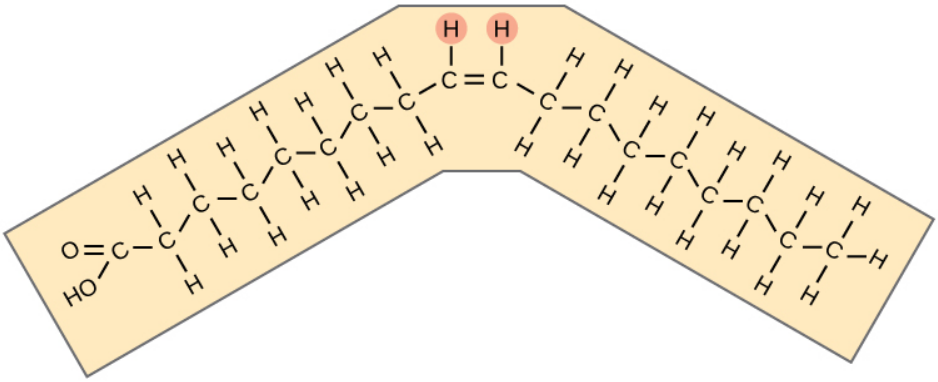
Saturated fatty acid

Stearic acid



Unsaturated fatty acids

Cis oleic acid



Trans oleic acid

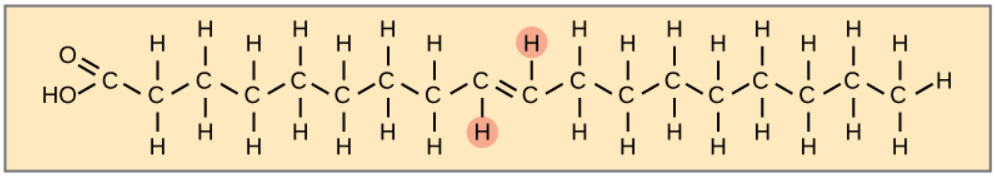


Figure 10.32 Saturated fatty acids have hydrocarbon chains connected by single bonds only. Unsaturated fatty acids have one or more double bonds. Each double bond may be in a *cis* or *trans* configuration. In the *cis* configuration, both hydrogens are on the same side of the hydrocarbon chain. In the *trans* configuration, the hydrogens are on opposite sides. A *cis* double bond causes a kink in the chain.

Trans Fats

In the food industry, oils are artificially hydrogenated to make them semi-solid and of a consistency desirable for many processed food products. Simply speaking, hydrogen gas is bubbled through oils to solidify them. During this hydrogenation process, double bonds of the *cis*- conformation in the hydrocarbon chain may be converted to double bonds in the trans-conformation.

Margarine, some types of peanut butter, and shortening are examples of artificially hydrogenated trans fats. Recent studies have shown that an increase in trans fats in the human diet may lead to an increase in levels of low-density lipoproteins (LDL), or “bad” cholesterol, which in turn may lead to plaque deposition in the arteries, resulting in heart disease. Many fast food restaurants have recently banned the use of trans fats, and food labels are required to display the trans fat content.

Omega Fatty Acids

Essential fatty acids are fatty acids required but not synthesized by the human body. Consequently, they have to be supplemented through ingestion via the diet. Omega-3 fatty acids (like that shown in **Figure 10.33**) fall into this category and are one of only two known for humans (the other being omega-6 fatty acid). These are polyunsaturated fatty acids and are called omega-3 because the third carbon from the end of the hydrocarbon chain is connected to its neighboring carbon by a double bond.

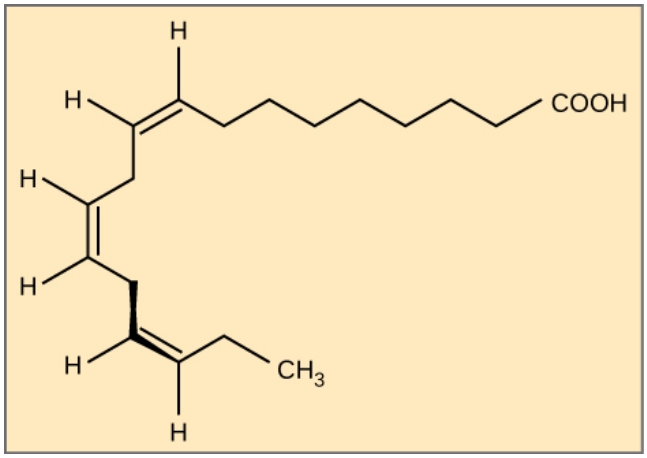


Figure 10.33 Alpha-linolenic acid is an example of an omega-3 fatty acid. It has three *cis* double bonds and, as a result, a curved shape. For clarity, the carbons are not shown. Each singly bonded carbon has two hydrogens associated with it, also not shown.

The farthest carbon away from the carboxyl group is numbered as the omega (ω) carbon, and if the double bond is between the third and fourth carbon from that end, it is known as an omega-3 fatty acid. Nutritionally important because the body does not make them, omega-3 fatty acids include alpha-linoleic acid (ALA), eicosapentaenoic acid (EPA), and docosahexaenoic acid (DHA), all of which are polyunsaturated. Salmon, trout, and tuna are good sources of omega-3 fatty acids. Research indicates that omega-3 fatty acids reduce the risk of sudden death from heart attacks, reduce triglycerides in the blood, lower blood pressure, and prevent thrombosis by inhibiting blood clotting. They also reduce inflammation, and may help reduce the risk of some cancers in animals.

Like carbohydrates, fats have received a lot of bad publicity. It is true that eating an excess of fried foods and other "fatty" foods leads to weight gain. However, fats do have important functions. Many vitamins are fat soluble, and fats serve as a long-term storage form of fatty acids: a source of energy. They also provide insulation for the body. Therefore, "healthy" fats in moderate amounts should be consumed on a regular basis.

Waxes

Wax covers the feathers of some aquatic birds and the leaf surfaces of some plants. Because of the hydrophobic nature of waxes, they prevent water from sticking on the surface (**Figure 10.34**). Waxes are made up of long fatty acid chains esterified to long-chain alcohols.

Figure 10.34 Waxy coverings on some leaves are made of lipids. (credit: Roger Griffith)

Phospholipids

Phospholipids are major constituents of the plasma membrane, the outermost layer of animal cells. Like fats, they are composed of fatty acid chains attached to a glycerol or sphingosine backbone. Instead of three fatty acids attached as in triglycerides, however, there are two fatty acids forming diacylglycerol, and the third carbon of the glycerol backbone is

occupied by a modified phosphate group (Figure 10.35). A phosphate group alone attached to a diaglycerol does not qualify as a phospholipid; it is phosphatidate (diacylglycerol 3-phosphate), the precursor of phospholipids. The phosphate group is modified by an alcohol. Phosphatidylcholine and phosphatidylserine are two important phospholipids that are found in plasma membranes.

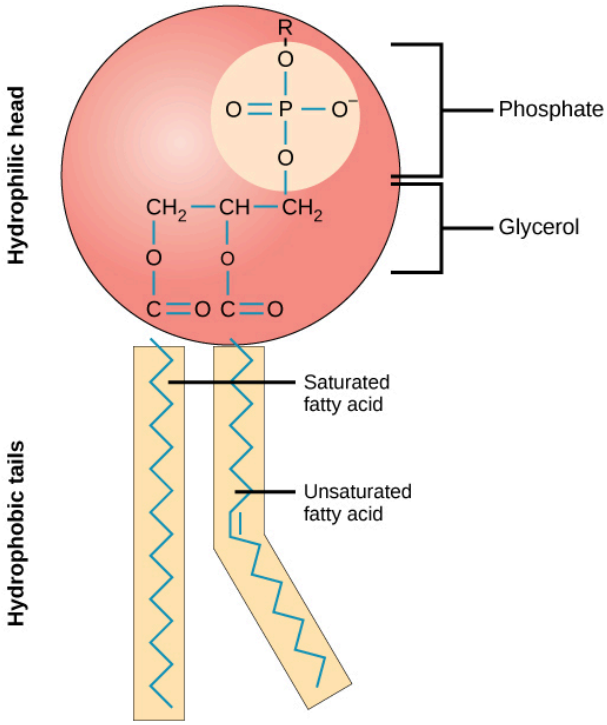


Figure 10.35 A phospholipid is a molecule with two fatty acids and a modified phosphate group attached to a glycerol backbone. The phosphate may be modified by the addition of charged or polar chemical groups. Two chemical groups that may modify the phosphate, choline and serine, are shown here. Both choline and serine attach to the phosphate group at the position labeled R via the hydroxyl group indicated in green.

A phospholipid is an amphipathic molecule, meaning it has a hydrophobic and a hydrophilic part. The fatty acid chains are hydrophobic and cannot interact with water, whereas the phosphate-containing group is hydrophilic and interacts with water (Figure 10.36).

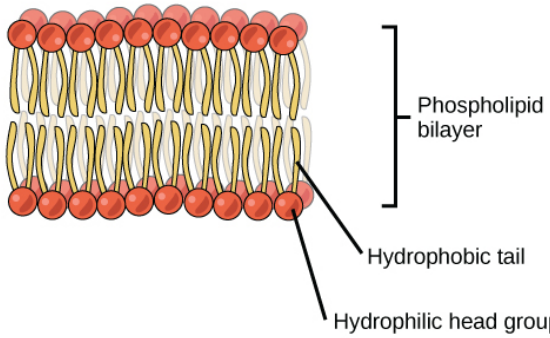


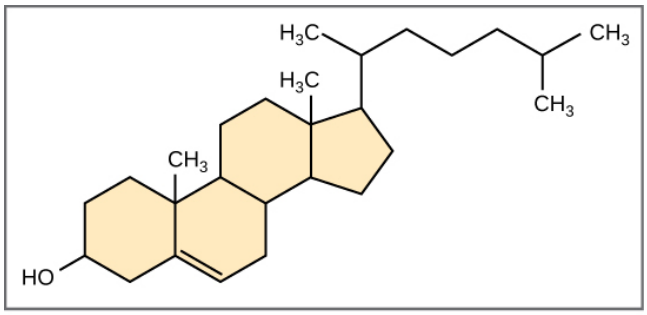
Figure 10.36 The phospholipid bilayer is the major component of all cellular membranes. The hydrophilic head groups of the phospholipids face the aqueous solution. The hydrophobic tails are sequestered in the middle of the bilayer.

The head is the hydrophilic part, and the tail contains the hydrophobic fatty acids. In a membrane, a bilayer of phospholipids forms the matrix of the structure, the fatty acid tails of phospholipids face inside, away from water, whereas the phosphate group faces the outside, aqueous side (Figure 10.36).

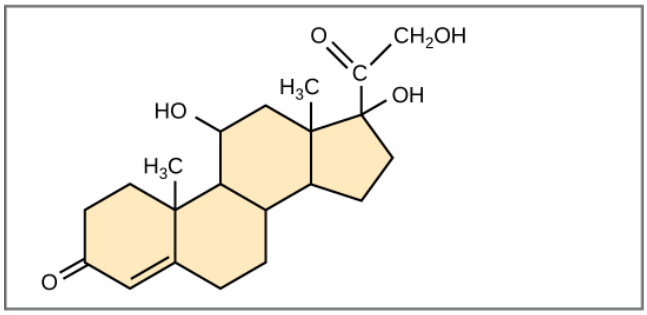
Phospholipids are responsible for the dynamic nature of the plasma membrane. If a drop of phospholipids is placed in water, it spontaneously forms a structure known as a micelle, where the hydrophilic phosphate heads face the outside and the fatty acids face the interior of this structure.

Steroids

Unlike the phospholipids and fats discussed earlier, **steroids** have a fused ring structure. Although they do not resemble the other lipids, they are grouped with them because they are also hydrophobic and insoluble in water. All steroids have four linked carbon rings and several of them, like cholesterol, have a short tail (**Figure 10.37**). Many steroids also have the –OH functional group, which puts them in the alcohol classification (sterols).



Cholesterol



Cortisol

Figure 10.37 Steroids such as cholesterol and cortisol are composed of four fused hydrocarbon rings.

Cholesterol is the most common steroid. Cholesterol is mainly synthesized in the liver and is the precursor to many steroid hormones such as testosterone and estradiol, which are secreted by the gonads and endocrine glands. It is also the precursor to Vitamin D. Cholesterol is also the precursor of bile salts, which help in the emulsification of fats and their subsequent absorption by cells. Although cholesterol is often spoken of in negative terms by lay people, it is necessary for proper functioning of the body. It is a component of the plasma membrane of animal cells and is found within the phospholipid bilayer. Being the outermost structure in animal cells, the plasma membrane is responsible for the transport of materials and cellular recognition and it is involved in cell-to-cell communication.

10.7 | Components and Structure of Plasmal Membranes

Introduction

> “The American Heritage Dictionary defines a membrane as a 'thin pliable layer of plant or animal tissue covering or separating structures or organs'. The impression this description leaves is one of the plastic wrap covering a hamburger. By this definition, membranes are static, tough, impenetrable, and visible. Yet,

> nothing could be further from the truth. The entire concept of **dynamic behavior** is missing from this definition, yet dynamics is what makes membranes both essential for life and so difficult to study.”

William Stillwell, *An Introduction to Biological Membranes: From Bilayers to Rafts*, pg. 1, 2013

A cell’s plasma membrane defines the cell, outlines its borders, and determines the nature of its interaction with its environment. As Stillwell says above, without membranes there would be no life; they are as essential to life as DNA or proteins. Cells exclude some substances, take in others, and excrete still others, all in controlled quantities. The plasma membrane must be very flexible to allow certain cells, such as red blood cells and white blood cells, to change shape as they pass through narrow capillaries. These are the more obvious functions of a plasma membrane. In addition, the surface of the plasma membrane carries markers that allow cells to recognize one another, which is vital for tissue and organ formation during early development, and which later plays a role in the “self” versus “non-self” distinction of the immune response.

Among the most sophisticated functions of the plasma membrane is the ability to transmit signals by means of complex, integral proteins known as **membrane receptors**. These proteins (and occasionally, lipids) act both as receivers of extracellular inputs and as activators of intracellular processes. These membrane receptors provide extracellular attachment sites for effectors like hormones and growth factors, and they activate intracellular response cascades when their effectors are bound. Occasionally, receptors are hijacked by viruses (HIV, human immunodeficiency virus, is one example) that use them to gain entry into cells, and at times, the genes encoding receptors become mutated, causing the process of signal transduction to malfunction with disastrous consequences.

Fluid Mosaic Model

The existence of the plasma membrane was identified in the 1890s, and its chemical components were identified in 1915. The principal components identified at that time were lipids and proteins. The first widely accepted model of the plasma membrane’s structure was proposed in 1935 by Hugh Davson and James Danielli; it was based on the “railroad track” appearance of the plasma membrane in early electron micrographs. They theorized that the structure of the plasma membrane resembles a sandwich, with protein being analogous to the bread, and lipids being analogous to the filling. In the 1950s, advances in microscopy, notably transmission electron microscopy (TEM), allowed researchers to see that the core of the plasma membrane consisted of a double, rather than a single, layer. A new model that better explains both the microscopic observations and the function of that plasma membrane was proposed by S.J. Singer and Garth L. Nicolson in 1972.

The explanation proposed by Singer and Nicolson, and based on the work of many others such as Harden McConnell, is called the **fluid mosaic model**. The model has evolved somewhat over time, but it still best accounts for the structure and functions of the plasma membrane as we now understand them. The fluid mosaic model describes the structure of the plasma membrane as a mosaic of components—including phospholipids, cholesterol, proteins, and carbohydrates—that gives the membrane a fluid character. Plasma membranes range from 5 to 10 nm in thickness. For comparison, human red blood cells, visible via light microscopy, are approximately 8 µm wide, or approximately 1,000 times wider than a plasma membrane. The membrane does look a bit like a sandwich (**Figure 10.38**).

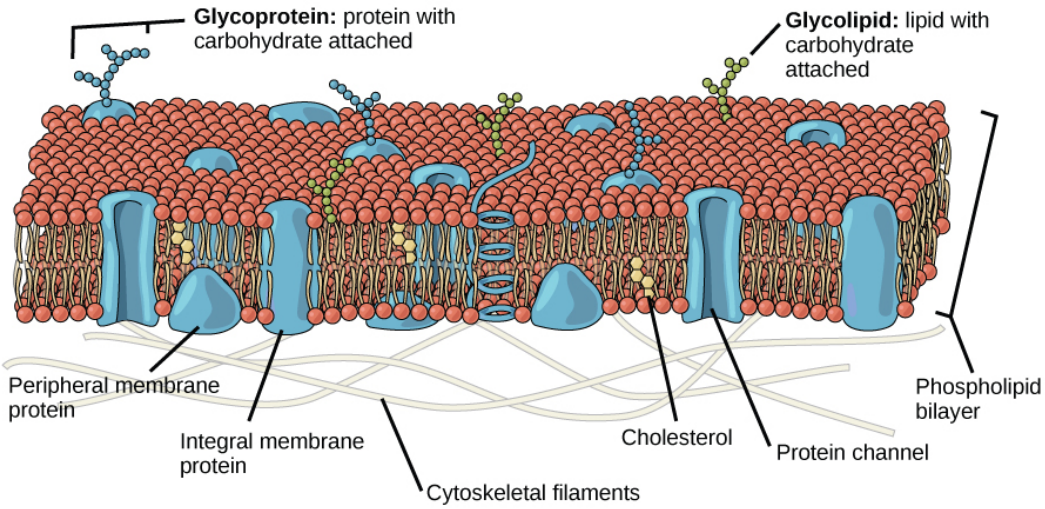


Figure 10.38 The fluid mosaic model of the plasma membrane describes the plasma membrane as a fluid combination of phospholipids, cholesterol, and proteins. Carbohydrates attached to lipids (glycolipids) and to proteins (glycoproteins) extend from the outward-facing surface of the membrane.

The principal components of a plasma membrane are lipids (phospholipids and cholesterol), proteins, and carbohydrates attached to some of the lipids and some of the proteins. A **phospholipid** is a molecule consisting of glycerol, two fatty acids, and a phosphate-linked head group. **Cholesterol**, another lipid composed of four fused carbon rings, is found alongside the phospholipids in the core of the membrane. The proportions of proteins, lipids, and carbohydrates in the plasma membrane vary with cell type, but for a typical human cell, protein accounts for about 50 percent of the composition by mass, lipids (of all types) account for about 40 percent of the composition by mass, with the remaining 10 percent of the composition by mass being carbohydrates. However, the concentration of proteins and lipids varies with different biological membranes. For example, myelin, an outgrowth of the membrane of specialized cells that insulates the axons of the peripheral nerves, contains only 18 percent protein and 76 percent lipid. The mitochondrial inner membrane contains 76 percent protein and only 24 percent lipid. The plasma membrane of human red blood cells is 30 percent lipid. Carbohydrates are present only on the exterior surface of the plasma membrane and are attached to proteins, forming glycoproteins, or attached to lipids, forming glycolipids.

Phospholipids

The main fabric of the membrane is composed of amphiphilic, phospholipid molecules. The **hydrophilic** or “water-loving” areas of these molecules (which look like a collection of balls in an artist’s rendition of the model) (**Figure 10.38**) are in contact with the aqueous fluid both inside and outside the cell. **Hydrophobic**, or water-hating molecules, tend to be non-polar. They interact with other non-polar molecules in chemical reactions, but generally do not interact with polar molecules. When placed in water, hydrophobic molecules tend to form a ball or cluster. The hydrophilic regions of the phospholipids tend to form hydrogen bonds with water and other polar molecules on both the exterior and interior of the cell. Thus, the membrane surfaces that face the interior and exterior of the cell are hydrophilic. In contrast, the interior of the plasma membrane is hydrophobic and will not interact with water. Therefore, phospholipids form an excellent two-layer biological membrane that separates fluid within the cell from the fluid outside of the cell.

A phospholipid molecule (**Figure 10.39**) consists of a three-carbon glycerol backbone with two fatty acid molecules attached to carbons 1 and 2, and a phosphate-containing group attached to the third carbon. This arrangement gives the overall molecule an area described as its head (the phosphate-containing group), which has a polar character or negative charge, and an area called the tail (the fatty acids), which has no charge. The head can form hydrogen bonds, but the tail cannot. A molecule with this arrangement of a positively or negatively charged area and an uncharged, or non-polar, area is referred to as **amphiphilic** or “dual-loving.”

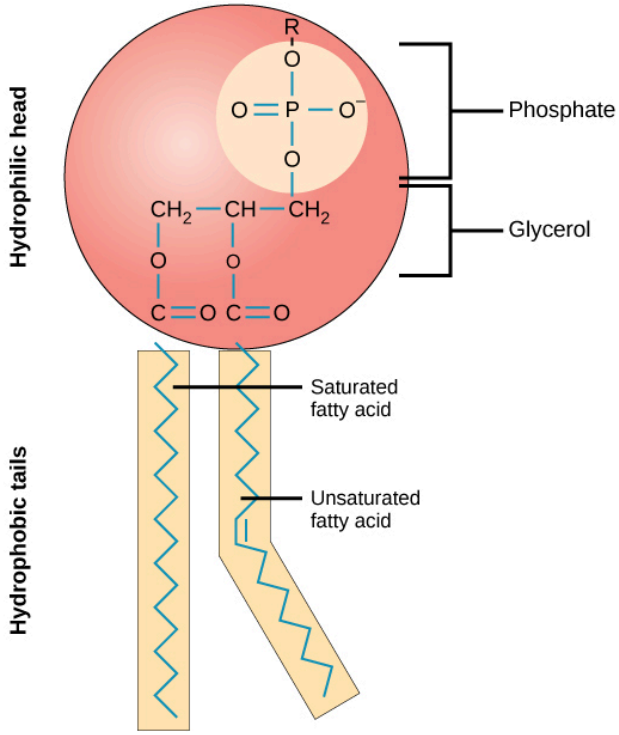


Figure 10.39 This phospholipid molecule is composed of a hydrophilic head and two hydrophobic tails. The hydrophilic head group consists of a phosphate-containing group attached to a glycerol molecule. The hydrophobic tails, each containing either a saturated or an unsaturated fatty acid, are long hydrocarbon chains.

This characteristic is vital to the structure of a plasma membrane because, in water, phospholipids tend to become arranged with their hydrophobic tails facing each other and their hydrophilic heads facing out. In this way, they form a lipid bilayer—a barrier composed of a double layer of phospholipids that separates the water and other materials on one side of the barrier from the water and other materials on the other side. In fact, phospholipids heated in an aqueous solution tend to spontaneously form small spheres or droplets (called micelles or liposomes), with their hydrophilic heads forming the exterior and their hydrophobic tails on the inside (**Figure 10.40**).

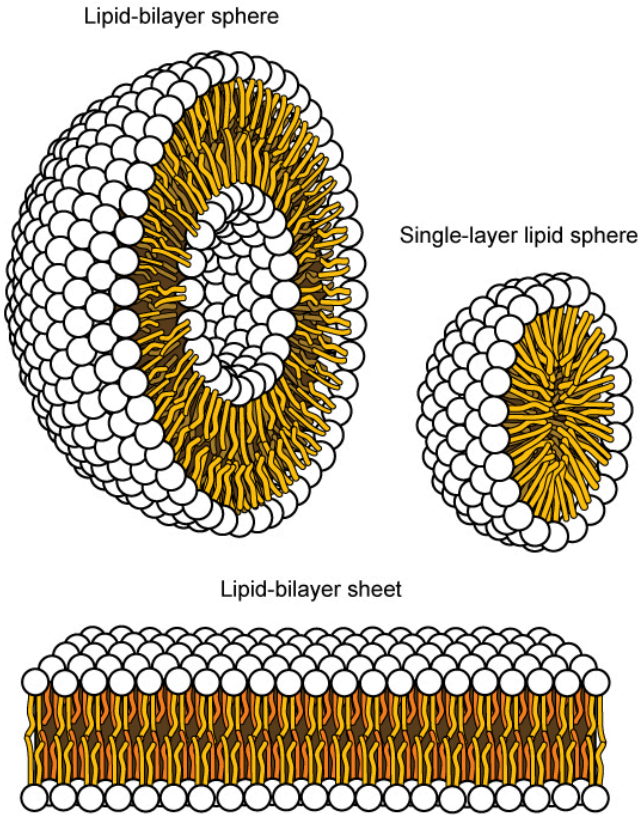


Figure 10.40 In an aqueous solution, phospholipids tend to arrange themselves with their polar heads facing outward and their hydrophobic tails facing inward. (credit: modification of work by Mariana Ruiz Villareal)

Proteins

Proteins make up the second major component of plasma membranes. Integral proteins (some specialized types are called integrins) are, as their name suggests, integrated completely into the membrane structure, and their hydrophobic membrane-spanning regions interact with the hydrophobic region of the the phospholipid bilayer (**Figure 10.38**). Single-pass integral membrane proteins usually have a hydrophobic transmembrane segment that consists of 20–25 amino acids. Some span only part of the membrane—associating with a single layer—while others stretch from one side of the membrane to the other, and are exposed on either side. Some complex proteins are composed of up to 12 segments of a single protein, which are extensively folded and embedded in the membrane (**Figure 10.41**). This type of protein has a hydrophilic region or regions, and one or several mildly hydrophobic regions. This arrangement of regions of the protein tends to orient the protein alongside the phospholipids, with the hydrophobic region of the protein adjacent to the tails of the phospholipids and the hydrophilic region or regions of the protein protruding from the membrane and in contact with the cytosol or extracellular fluid.

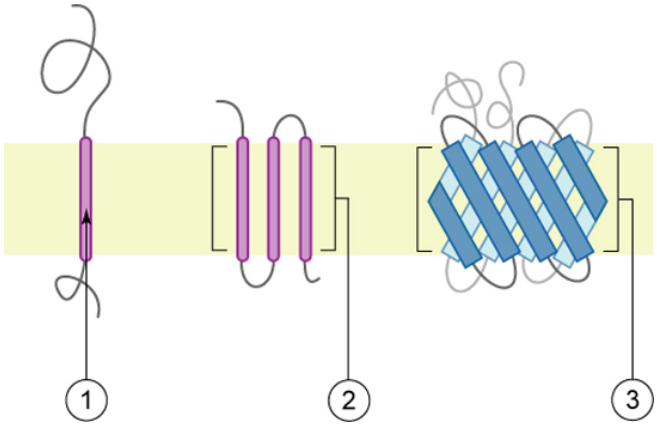


Figure 10.41 Integral membranes proteins may have one or more alpha-helices that span the membrane (examples 1 and 2), or they may have beta-sheets that span the membrane (example 3). (credit: “Foobar”/Wikimedia Commons)

Peripheral proteins are found on the exterior and interior surfaces of membranes, attached either to integral proteins or to phospholipids. Peripheral proteins, along with integral proteins, may serve as enzymes, as structural attachments for the fibers of the cytoskeleton, or as part of the cell's recognition sites. These are sometimes referred to as "cell-specific" proteins. The body recognizes its own proteins and attacks foreign proteins associated with invasive pathogens.

11 | TOUR OF THE CELL: PROTEINS AND ENZYME FUNCTION

11.1 | Proteins

Introduction

> "The characteristic specific properties of native proteins we attribute to their uniquely defined configurations. The denatured protein molecule we consider to be characterized by the absence of a uniquely defined configuration."

Alfred E. Mirsky and Linus Pauling, "On the Structure of Native, Denatured and Coagulated Proteins", *Proceedings of the National Academy of Sciences of the United States of America*, 22:442-3, 1936

Proteins are one of the most abundant organic molecules in living systems and have the most diverse range of functions of all macromolecules. That diversity of function is due to a tremendous diversity of "uniquely defined" structures. Proteins may be structural, regulatory, contractile, or protective; they may serve in transport, storage, or membranes; or they may be toxins or enzymes. Each cell in a living system may contain thousands of proteins, each with a unique function. Their structures, like their functions, vary greatly. They are all, however, polymers of amino acids, arranged in a linear sequence. But that simple linear sequence is just the beginning of the story.

Types and Functions of Proteins

The primary types and functions of proteins are listed in **Table 11.1**. We will consider some of these categories in some detail, but the others will be left for later discussion.

Enzymes, which are produced by living cells, are **catalysts** in biochemical reactions (like digestion) and are usually complex or conjugated proteins. Each enzyme is specific for the substrate (a reactant that binds to an enzyme) it acts on. The enzyme may help in breakdown, rearrangement, or synthesis reactions. Enzymes that break down their substrates are called **catabolic** enzymes, and often this is a **hydrolysis** reaction. Enzymes that build more complex molecules from their substrates are called **anabolic** enzymes, and often this is **condensation reaction** or dehydration synthesis. It should be noted that all enzymes increase the rate of reaction and, therefore, are considered to be organic catalysts. An example of an enzyme is salivary amylase, which hydrolyzes (breaks down) its substrate amylose, a component of starch, producing the simple disaccharide known as maltose along with other simpler sugars. Lastly, most but not all enzymes are proteins. Some enzymes are composed of RNA (Ribonucleic Acid) or have RNA components; a topic covered in future chapters.

Hormones are chemical-signaling molecules, usually small proteins or steroids, secreted by endocrine cells that act to control or regulate specific physiological processes, including growth, development, metabolism, and reproduction. For example, insulin is a protein hormone that helps to regulate the blood glucose level.

Structural proteins are some of the more familiar proteins encountered everyday. Hair, fingernails, and feathers are largely composed of proteins called keratins. Your skin contains large quantities of proteins called collagens and elastins. Other structural proteins are found in bone, in muscle, in connective tissue, etc.

Storage proteins are used by some organisms to store energy over the long term, just as carbohydrates and lipids are the preferred energy storage molecules for other organisms. Casein, a protein found in milk, is one example. Zein proteins

found in wheat grains provide energy for the developing wheat embryo, but also are critical in helping bread dough to rise and hold its shape. Egg albumin is an energy source for bird embryonic development. And proteins found in legumes, such as soybeans and other beans, nourish the embryos of those plants, as well as billions of humans around the world.

Protein Types and Functions

Type	Examples	Functions
Catabolic/ Digestive Enzymes	Amylase, lipase, pepsin, trypsin	Help in digestion of food by catabolizing nutrients into monomers
Anabolic Enzymes	DNA polymerase, glycogen synthase	Enzymes that make polymers from monomers
Transport	Hemoglobin, albumin	Carry substances in the blood or lymph throughout the body
Structural	Actin, tubulin, keratin, collagen	Compose structures that support cell organelles (e.g. cytoskeleton) or body parts (e.g. tendons, cartilage)
Hormones	Insulin, thyroxine	Coordinate the activity of different body systems
Defense	Immunoglobulins	Protect the body from foreign pathogens
Contractile	Actin, myosin	Muscle contraction
Storage	Legume storage proteins, egg white (albumin)	Provide nourishment in early development of the embryo and the seedling

Table 11.1

Proteins have different shapes and molecular weights; some proteins are globular in shape whereas others are fibrous in nature. For example, hemoglobin is a globular protein, but collagen, found in our skin, is a fibrous protein. Protein shape is critical to its function, and this shape is maintained by many different types of chemical bonds. Changes in temperature, pH, salinity and exposure to chemicals may lead to permanent changes in the shape of the protein, leading to loss of function, known as **denaturation**. All proteins are made up of different arrangements of the same 20 types of amino acids.

Amino Acids

Amino acids are the monomers that make up proteins. Each amino acid has the same fundamental structure, which consists of a central carbon atom, also known as the alpha (α) carbon, bonded to an amino group (NH_2), a carboxyl group (COOH), and to a hydrogen atom. Every amino acid also has another atom or group of atoms bonded to the central atom known as the R group (Figure 11.1).

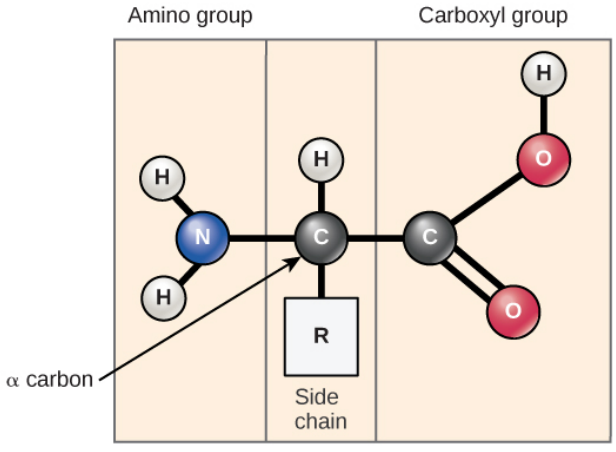


Figure 11.1 Amino acids have a central asymmetric carbon to which an amino group, a carboxyl group, a hydrogen atom, and a side chain (R group) are attached.

The name "amino acid" is derived from the fact that they contain both amino group and carboxyl-acid-group in their basic structure. As mentioned, there are 20 amino acids present in proteins. Ten of these are considered essential amino acids in humans because the human body cannot produce them and they are obtained from the diet. For each amino acid, the R group (or side chain) is different (**Figure 11.2**).

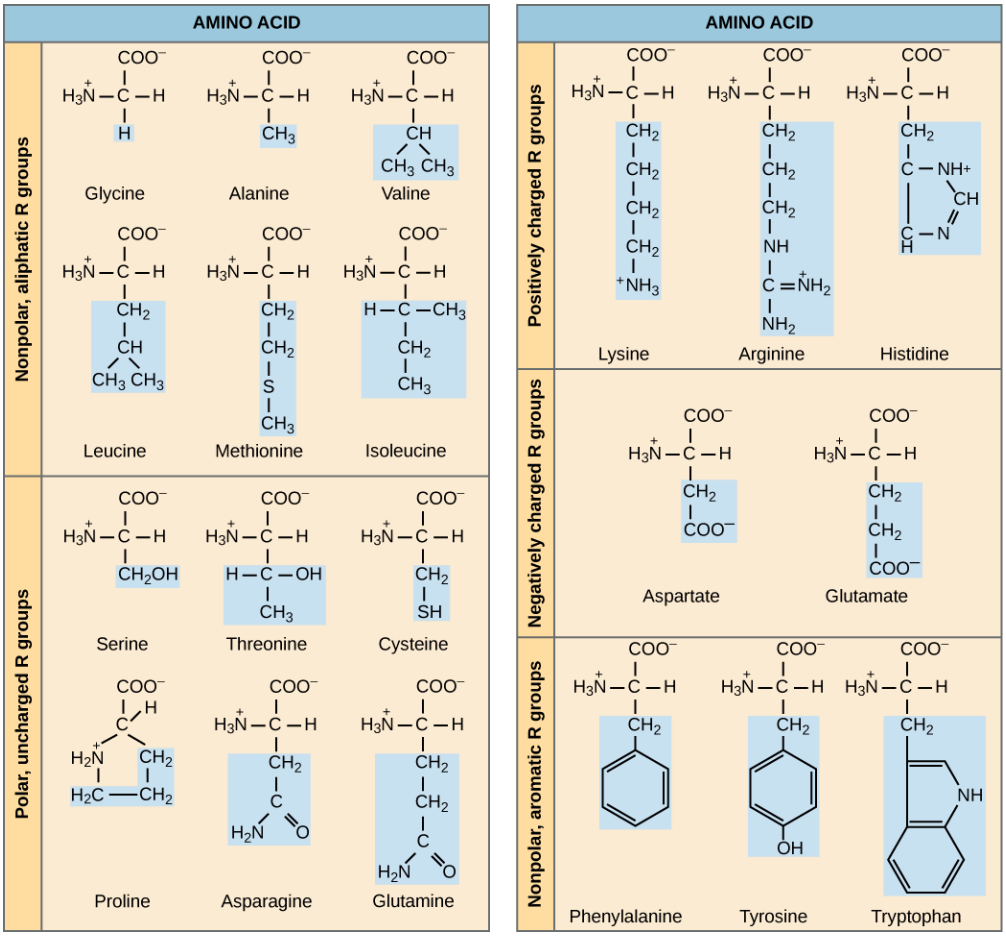


Figure 11.2 There are 20 common amino acids commonly found in proteins, each with a different R group (variant group) that determines its chemical nature.

The chemical nature of the side chain determines the nature of the amino acid (that is, whether it is acidic, basic, polar, or nonpolar). For example, the amino acid glycine has a hydrogen atom as the R group. Amino acids such as valine, methionine, and alanine are nonpolar or hydrophobic in nature, while amino acids such as serine, threonine, and cysteine are polar and have hydrophilic side chains. The side chains of lysine and arginine are positively charged, and therefore these amino acids are also known as basic amino acids. Proline has an R group that is linked to the amino group, forming a ring-like structure. Proline is an exception to the standard structure of an amino acid since its amino group is not separate from the side chain (**Figure 11.2**).

Amino acids are represented by a single upper case letter or a three-letter abbreviation. For example, valine is known by the letter V or the three-letter symbol val. Just as some fatty acids are essential to a diet, some amino acids are necessary as well. They are known as essential amino acids, and in humans they include isoleucine, leucine, and cysteine. Essential amino acids refer to those necessary for construction of proteins in the body, although not produced by the body; which amino acids are essential varies from organism to organism.

The sequence and the number of amino acids ultimately determine the protein's shape, size, and function. Each amino acid is attached to another amino acid by a covalent bond, known as a **peptide bond**, which is formed by a condensation reaction. The carboxyl group of one amino acid and the amino group of the incoming amino acid combine, releasing a molecule of water. The resulting bond is the peptide bond (**Figure 11.3**).

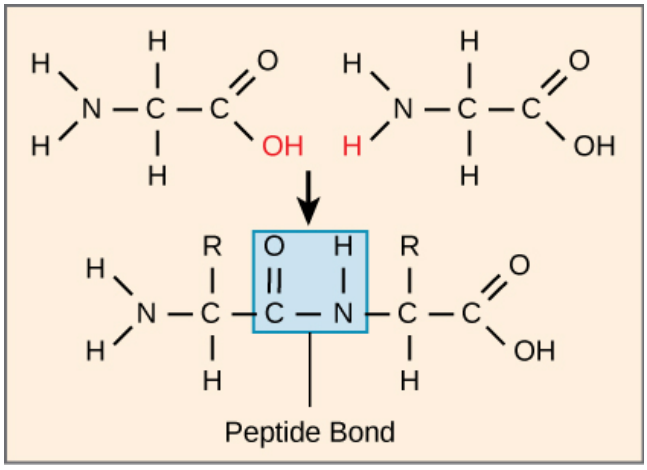


Figure 11.3 Peptide bond formation is a condensation reaction. The carboxyl group of one amino acid is linked to the amino group of the incoming amino acid. In the process, a molecule of water is released.

The products formed by such linkages are called **peptides**. As more amino acids join to this growing chain, the resulting chain is known as a **polypeptide**. Each polypeptide has a free amino group at one end. This end is called the N terminal, or the amino terminal, and the other end has a free carboxyl group, also known as the C or carboxyl terminal. While the terms polypeptide and protein are sometimes used interchangeably, a polypeptide is technically a polymer of amino acids, whereas the term protein is used for a polypeptide or polypeptides that have combined together, often have bound non-peptide prosthetic groups, have a distinct shape, and have a unique function. After protein synthesis (translation), most proteins are modified. These are known as post-translational modifications. They may undergo cleavage, phosphorylation, or may require the addition of other chemical groups. Only after these modifications is the protein completely functional.

evolution CONNECTION

The Evolutionary Significance of Cytochrome c

Cytochrome c is an important component of the electron transport chain, a part of cellular respiration, and it is normally found in the cellular organelle, the mitochondrion. This protein has a heme prosthetic group, and the central ion of the heme gets alternately reduced and oxidized during electron transfer. Because this essential protein's role in producing cellular energy is crucial, it has changed very little over millions of years. Protein sequencing has shown that there is a considerable amount of cytochrome c amino acid sequence homology among different species; in other words, evolutionary kinship can be assessed by measuring the similarities or differences among various species' DNA or protein sequences.

Scientists have determined that human cytochrome c contains 104 amino acids. For each cytochrome c molecule from different organisms that has been sequenced to date, 37 of these amino acids appear in the same position in all samples of cytochrome c. This indicates that there may have been a common ancestor. On comparing the human and chimpanzee protein sequences, no sequence difference was found. When human and rhesus monkey sequences were compared, the single difference found was in one amino acid. In another comparison, human to yeast sequencing shows a difference in the 44th position.

Protein Structure

As discussed earlier, the shape of a protein is critical to its function. For example, an enzyme can bind to a specific substrate at a site known as the active site. If this active site is altered because of local changes or changes in overall protein structure, the enzyme may be unable to bind to the substrate. To understand how the protein gets its final shape or conformation, we need to understand the four levels of protein structure: primary, secondary, tertiary, and quaternary.

Primary Structure

The unique sequence of amino acids in a polypeptide chain is its **primary structure**. For example, the pancreatic hormone insulin has two polypeptide chains, A and B, and they are linked together by disulfide bonds. The N terminal amino acid of the A chain is glycine, whereas the C terminal amino acid is asparagine (**Figure 11.4**). The sequences of amino acids in the

A and B chains are unique to insulin.

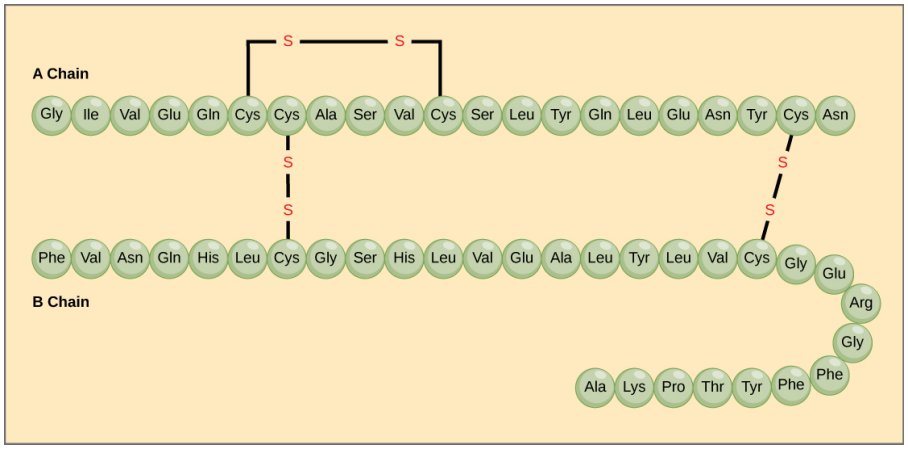


Figure 11.4 Bovine serum insulin is a protein hormone made of two peptide chains, A (21 amino acids long) and B (30 amino acids long). In each chain, primary structure is indicated by three-letter abbreviations that represent the names of the amino acids in the order they are present. The amino acid cysteine (cys) has a sulfhydryl (SH) group as a side chain. Two sulfhydryl groups can react in the presence of oxygen to form a disulfide (S-S) bond. Two disulfide bonds connect the A and B chains together, and a third helps the A chain fold into the correct shape. Note that all disulfide bonds are the same length, but are drawn different sizes for clarity.

The unique sequence for every protein is ultimately determined by the gene encoding the protein. A change in nucleotide sequence of the gene's coding region may lead to a different amino acid being added to the growing polypeptide chain, causing a change in protein structure and function. In sickle cell anemia, the hemoglobin β chain (a small portion of which is shown in **Figure 11.5**) has a single amino acid substitution, causing a change in protein structure and function. Specifically, the amino acid glutamic acid is substituted by valine in the β chain. What is most remarkable to consider is that a hemoglobin molecule is made up of two alpha chains and two beta chains that each consist of about 150 amino acids. The molecule, therefore, has about 600 amino acids. The structural difference between a normal hemoglobin molecule and a sickle cell molecule—which dramatically decreases life expectancy—is a single amino acid of the 600. What is even more remarkable is that those 600 amino acids are encoded by three nucleotides each, and the mutation is caused by a single base change (point mutation), 1 in 1800 bases.

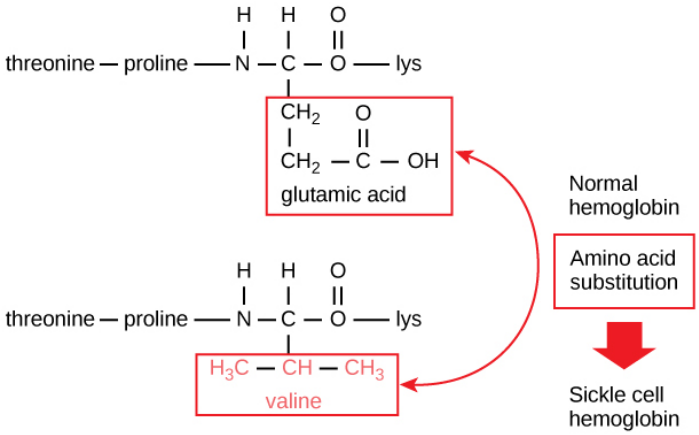


Figure 11.5 The beta chain of hemoglobin is 147 residues in length, yet a single amino acid substitution leads to sickle cell anemia. In normal hemoglobin, the amino acid at position seven is glutamate. In sickle cell hemoglobin, this glutamate is replaced by a valine.

Because of this change of one amino acid in the chain, hemoglobin molecules form long fibers that distort the biconcave, or disc-shaped, red blood cells and assume a crescent or "sickle" shape, which clogs arteries (**Figure 11.6**). This can lead to myriad serious health problems such as breathlessness, dizziness, headaches, and abdominal pain for those affected by this disease.

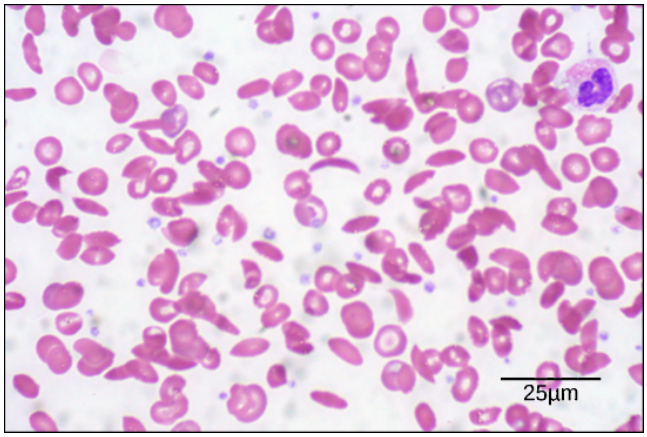


Figure 11.6 In this blood smear, visualized at 535x magnification using bright field microscopy, sickle cells are crescent shaped, while normal cells are disc-shaped. (credit: modification of work by Ed Uthman; scale-bar data from Matt Russell)

Secondary Structure

The local folding of the polypeptide in some regions gives rise to the **secondary structure** of the protein. The most common are the ***α*-helix** and ***β*-pleated sheet** structures (**Figure 11.7**). Both structure types are held in shape by hydrogen bonds. The hydrogen bonds form between the oxygen atom in the carbonyl group in one amino acid and another amino acid that is four amino acids farther along the chain.

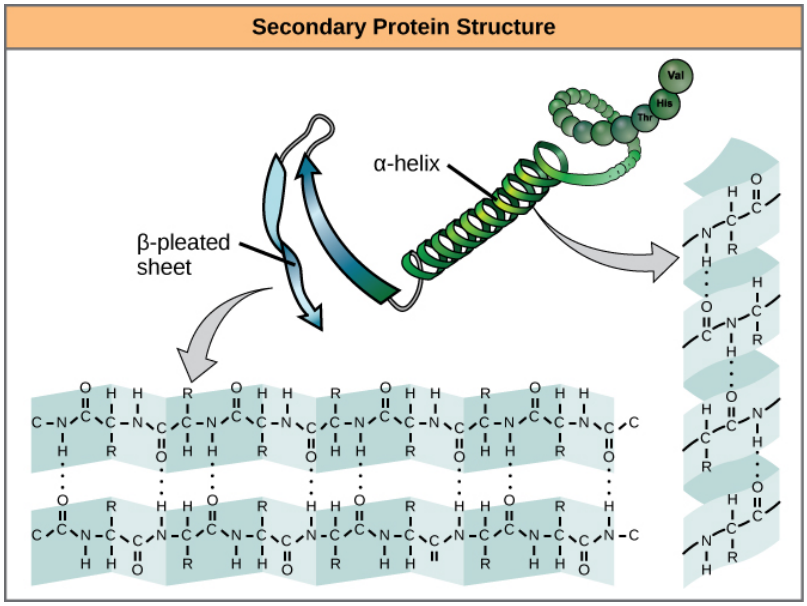


Figure 11.7 The *α*-helix and *β*-pleated sheet are secondary structures of proteins that form because of hydrogen bonding between carbonyl and amino groups in the peptide backbone. Certain amino acids have a propensity to form an *α*-helix, while others have a propensity to form a *β*-pleated sheet.

Every helical turn in an alpha helix has 3.6 amino acid residues. The R groups (the variant groups) of the polypeptide protrude out from the *α*-helix chain. In the *β*-pleated sheet, the "pleats" are formed by hydrogen bonding between atoms on the backbone of the polypeptide chain. The R groups are attached to the carbons and extend above and below the folds of the pleat. The pleated segments align parallel or antiparallel to each other, and hydrogen bonds form between the partially positive nitrogen atom in the amino group and the partially negative oxygen atom in the carbonyl group of the peptide backbone. The *α*-helix and *β*-pleated sheet structures are found in most globular and fibrous proteins and they play an important structural role.

Tertiary Structure

The unique three-dimensional structure of a polypeptide is its **tertiary structure** (**Figure 11.8**). This structure is in part due to chemical interactions at work on the polypeptide chain. Primarily, the interactions among R groups creates the

complex three-dimensional tertiary structure of a protein. The nature of the R groups found in the amino acids involved can counteract the formation of the hydrogen bonds described for standard secondary structures. For example, R groups with like charges are repelled by each other and those with unlike charges are attracted to each other (ionic bonds). When protein folding takes place, the hydrophobic R groups of nonpolar amino acids lay in the interior of the protein, whereas the hydrophilic R groups lay on the outside. The former types of interactions are also known as hydrophobic interactions. Interaction between cysteine side chains forms disulfide linkages in the presence of oxygen, the only covalent bond forming during protein folding.

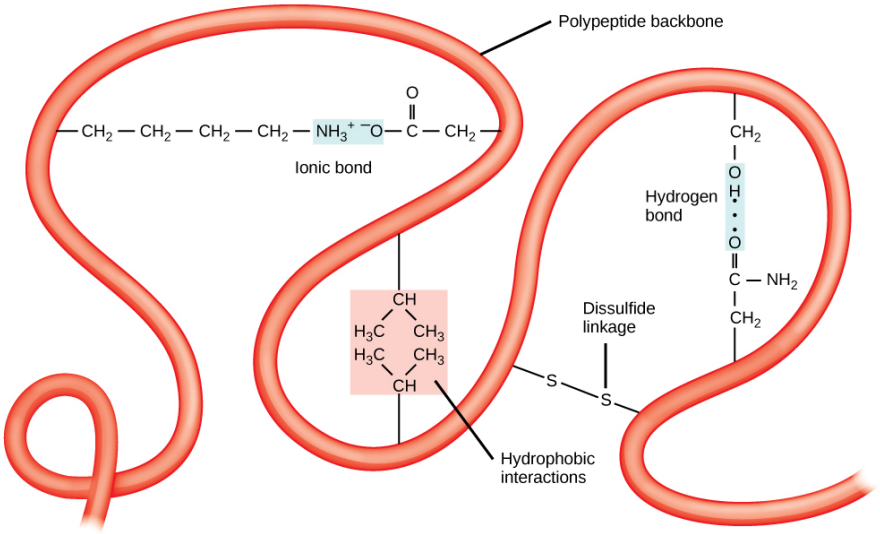


Figure 11.8 The tertiary structure of proteins is determined by a variety of chemical interactions. These include hydrophobic interactions, ionic bonding, hydrogen bonding and disulfide linkages.

All of these interactions, weak and strong, determine the final three-dimensional shape of the protein. When a protein loses its three-dimensional shape, it may no longer be functional.

Quaternary Structure

In nature, some proteins are formed from several polypeptides, also known as subunits, and the interaction of these subunits forms the **quaternary structure**. Weak interactions between the subunits help to stabilize the overall structure. For example, insulin (a globular protein) has a combination of hydrogen bonds and disulfide bonds that cause it to be mostly clumped into a ball shape. Insulin starts out as a single polypeptide and loses some internal sequences in the presence of post-translational modification after the formation of the disulfide linkages that hold the remaining chains together. Silk (a fibrous protein), however, has a β-pleated sheet structure that is the result of hydrogen bonding between different chains.

The four levels of protein structure (primary, secondary, tertiary, and quaternary) are illustrated in **Figure 11.9**.

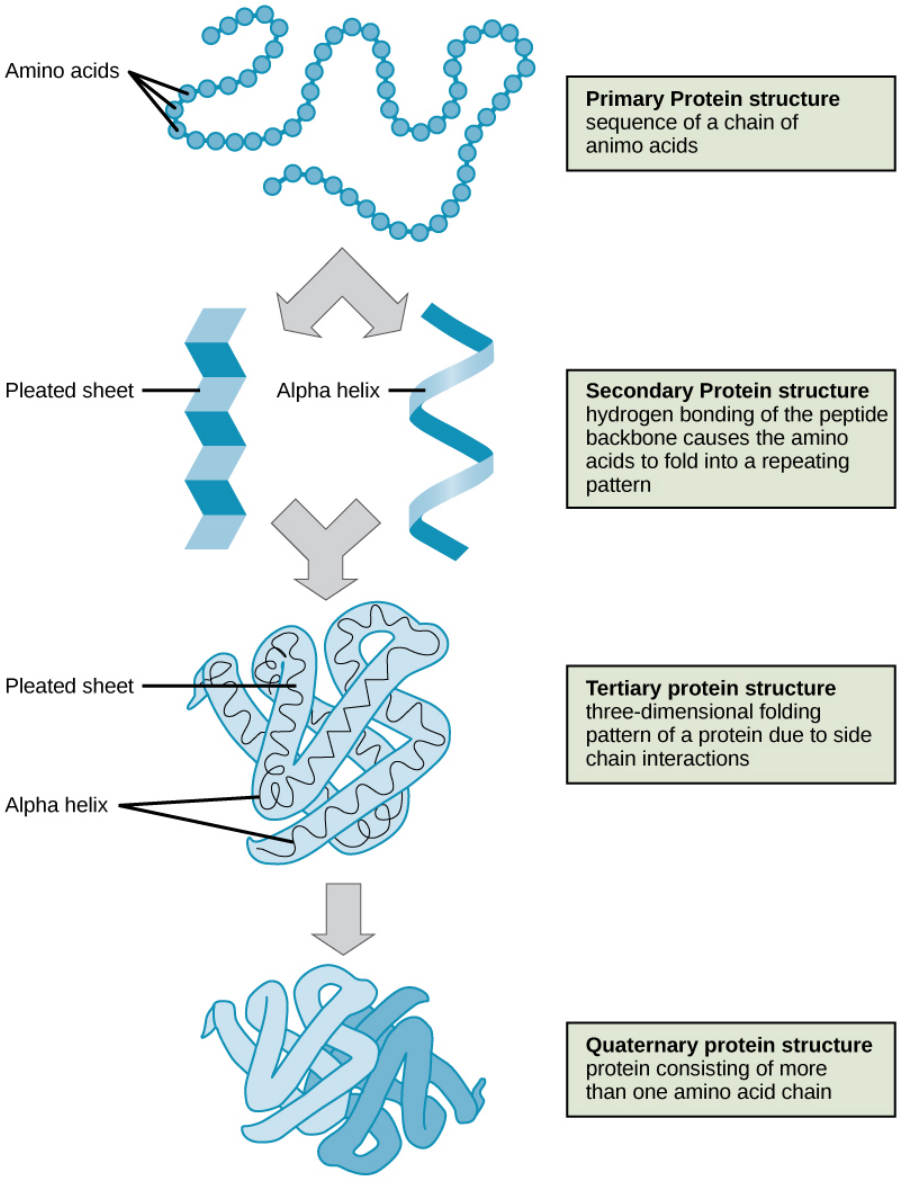


Figure 11.9 The four levels of protein structure can be observed in these illustrations. (credit: modification of work by National Human Genome Research Institute)

Denaturation and Protein Folding

Each protein has its own unique sequence and shape that are held together by chemical interactions (covalent, ionic and hydrogen bonds). As noted by Mirsky and Pauling in the epigraph above, a denatured protein is one that has lost that unique shape and configuration. If the protein is subject to changes in temperature, pH, salinity or exposure to chemicals, the protein structure may change, losing its shape without losing its primary sequence in what is known as **denaturation**. During denaturation, the changes in the environment surrounding the protein alter the chemical interactions (ionic and hydrogen bonds) within the protein causing changes in the shape and configuration of that protein. Denaturation may be reversible because the primary structure of the polypeptide is conserved in the process if the denaturing agent is removed, allowing the protein to resume its function. However, denaturation is often irreversible, leading to loss of function. One example of irreversible protein denaturation is when an egg is fried. The albumin protein in the liquid egg white is denatured when placed in a hot pan. Not all proteins are denatured at high temperatures; for instance, bacteria that survive in hot springs have proteins that function at temperatures close to boiling. The stomach is also very acidic, has a low pH, and denatures proteins as part of the digestion process; however, the digestive enzymes of the stomach retain their activity under these conditions.

Protein folding is critical to its function. It was originally thought that the proteins themselves were responsible for the

folding process. Only recently was it found that often they receive assistance in the folding process from protein helpers known as chaperones (or chaperonins) that associate with the target protein during the folding process. They act by preventing aggregation of polypeptides that make up the complete protein structure, and they disassociate from the protein once the target protein is folded.

11.2 | Enzymes

> *"Yet even after Galileo and Newton, there remained another question: Were living things somehow different from rocks and water and stars? Did animate and inanimate matter differ in some fundamental way? the "vitalists" claimed that animate matter had some special essence, an intangible spirit or soul, while the "mechanists" argued that living things were elaborate machines and obeyed precisely the same laws of physics and chemistry as did inanimate material."*

Alan Lightman, "Our Place in the Universe", *Harper's Magazine*, December 2012.

One of the greatest scientists of all time, Louis Pasteur, believed that metabolic reactions such as fermentation could **only** occur in living cells. This perspective was widely shared by scientists of the 19th Century. Thus one of the first damaging blows to the vitalist perspective came in the late 19th Century, when Eduard Buchner showed that a cell-free extract from yeast could carry out the synthesis of ethanol from glucose (alcoholic fermentation). Whole cells were not required for this process. These extracts were called " **enzymes**", which derives from the Greek words "en" (in) + "zyme" (yeast). Buchner demonstrated that there was something "in yeast", as opposed to the yeast cells themselves, that could convert glucose to ethanol. The subsequent work of many other scientists has built on this simple concept - cells contain substances that catalyze reactions, and these substances can be studied apart from the cells themselves.

A substance that helps a chemical reaction to occur is a **catalyst**, and the special molecules that catalyze biochemical reactions are called enzymes. Almost all enzymes are proteins, made up of chains of amino acids, and they perform the critical task of lowering the activation energies of chemical reactions inside the cell. Enzymes do this by binding to the reactant molecules, and holding them in such a way as to make the chemical bond-breaking and bond-forming processes take place more readily. It is important to remember that enzymes don’t change the ΔG of a reaction. In other words, they don’t change whether a reaction is exergonic (spontaneous) or endergonic. This is because they don’t change the free energy of the reactants or products. They only reduce the activation energy required to reach the transition state (**Figure 11.10**).

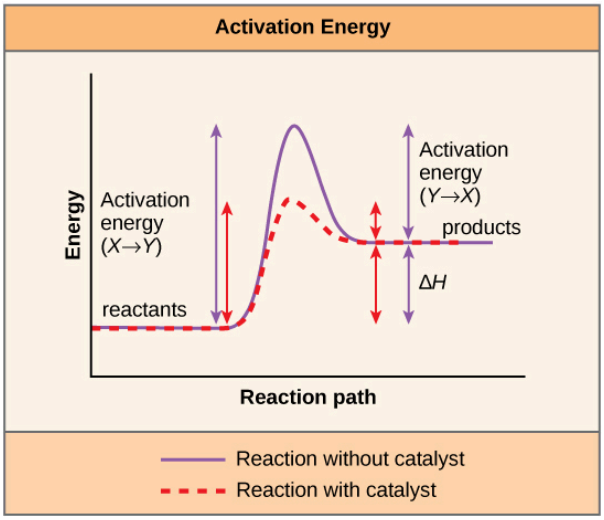


Figure 11.10 Enzymes lower the activation energy of the reaction but do not change the free energy of the reaction.

Enzyme Active Site and Substrate Specificity

The chemical reactants to which an enzyme binds are 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, both become modified, and 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, so to speak. Since enzymes are proteins, there is a unique combination of amino acid residues (also called side chains, or R groups) within the active site. Each residue is characterized by different properties. Residues can be large or small, weakly acidic or basic, hydrophilic or hydrophobic, positively or negatively charged, or neutral. The unique combination of amino acids, their positions, sequences, structures, and properties, creates a very specific chemical environment within the active site. This specific environment is suited to bind, albeit briefly, to a specific chemical substrate (or substrates). Due to this jigsaw puzzle-like match between an enzyme and its substrates (which adapts to find the best fit between the transition state and the active site), enzymes are known for their specificity. The "best fit" results from the shape and the amino acid functional group's attraction to the substrate. There is a specifically matched enzyme for each substrate and, thus, for each chemical reaction; however, there is flexibility as well.

Environmental factors influence enzyme activity

The fact that active sites are so perfectly suited to provide specific environmental conditions also means that they are subject to influences by the local environment. The environmental conditions can include temperature, pH, salinity and presence of heavy metals. It is true that increasing the environmental temperature generally increases reaction rates, enzyme-catalyzed or otherwise. However, increasing or decreasing the temperature outside of an optimal range can affect chemical bonds within the active site in such a way that they are less well suited to bind substrates. High temperatures will eventually cause enzymes, like other biological molecules, to **denature**, a process that changes the natural properties of a substance. Likewise, the pH of the local environment can also affect enzyme function. Active site amino acids have their own acidic or basic properties that are optimal for catalysis. These amino acids are sensitive to changes in pH that can impair the way substrate molecules bind. Enzymes are suited to function best within a certain pH range, and, as with temperature, extreme pH values (acidic or basic) of the environment can cause enzymes to denature an the rates of the reaction decrease (**Figure 11.11**).

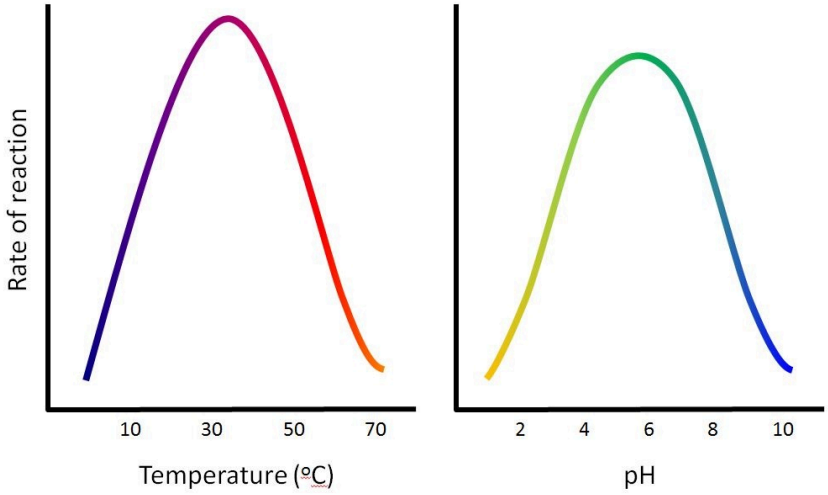


Figure 11.11 The effects of temperature and pH on the rate of enzyme catalyzed reaction.

Induced Fit and Enzyme Function

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 more refined view called **induced fit** (**Figure 11.12**). The induced-fit model expands upon the lock-and-key model by describing a more dynamic interaction between enzyme and substrate. As the enzyme and substrate come together, their interaction causes a mild shift in the enzyme's structure that confirms an ideal binding arrangement between the enzyme and the transition state of the substrate. This ideal binding maximizes the enzyme's ability to catalyze its reaction.

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 many ways. On a basic level, enzymes promote chemical reactions that involve more than one substrate by bringing the substrates together in an optimal orientation. The appropriate region (atoms and bonds) of one molecule is juxtaposed to the appropriate region of the other molecule with which it must react. 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. Certain chemical reactions might proceed best in a slightly acidic or non-polar environment. The chemical properties that emerge from the particular arrangement of amino acid residues within an active site create the perfect environment for an enzyme's specific substrates to react.

You've learned that the activation energy required for many reactions includes the energy involved in manipulating or slightly contorting chemical bonds so that they can easily break and allow others to reform. Enzymatic action can aid this process. The enzyme-substrate complex can lower the activation energy by contorting substrate molecules in such a way as to facilitate bond-breaking, helping to reach the transition state. Finally, enzymes can also lower activation energies by taking part in the chemical reaction itself. The amino acid residues can provide certain ions or chemical groups that actually form covalent bonds with substrate molecules as a necessary step of the reaction process. In these cases, it is important to remember that the enzyme will always return to its original state at 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 is done catalyzing a reaction, it releases its product(s).

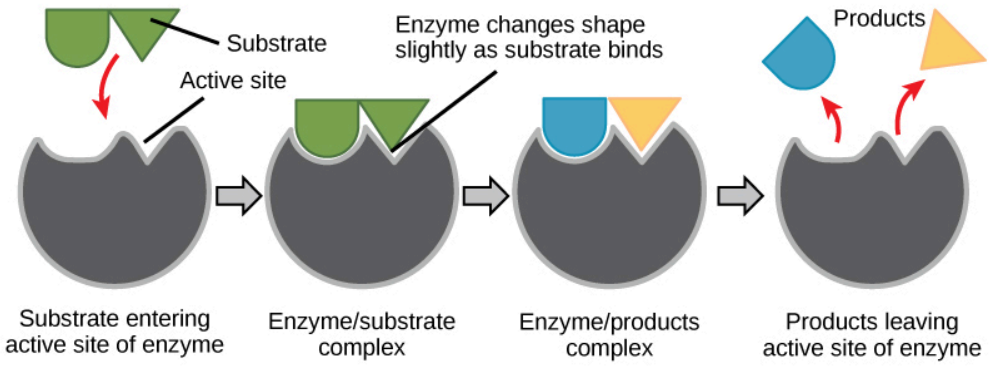


Figure 11.12 According to the induced-fit model, both enzyme and substrate undergo dynamic conformational changes upon binding. The enzyme contorts the substrate into its transition state, thereby increasing the rate of the reaction.

11.3 | Buffers and Enzymes

pH and Buffers

The pH of a solution indicates its acidity or alkalinity. The **pH scale** is, as previously mentioned, an inverse logarithm and ranges from 0 to 14 (**Figure 11.13**). Anything below 7.0 (ranging from 0.0 to 6.9) is acidic, and anything above 7.0 (from 7.1 to 14.0) is alkaline. Extremes in pH in either direction from 7.0 are usually considered inhospitable to life. The pH inside cells (6.8) and the pH in the blood (7.4) are both very close to neutral.

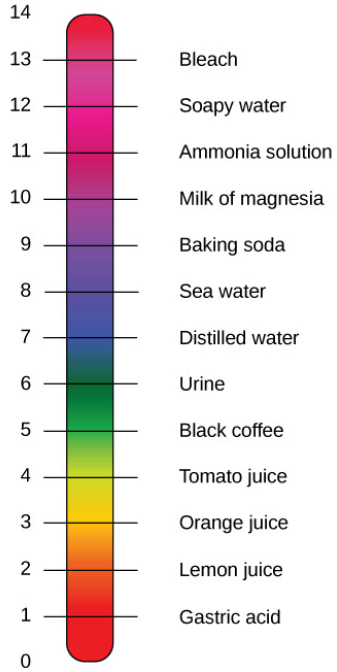


Figure 11.13 The pH scale measures the concentration of hydrogen ions (H^+) in a solution. (credit: modification of work by Edward Stevens)

So how can organisms whose bodies require a near-neutral pH ingest acidic and basic substances (a human drinking orange juice, for example) and survive? Buffers are the key. **Buffers** readily absorb excess H^+ or OH^-, keeping the pH of the body carefully maintained in the narrow range required for survival. Maintaining a constant blood pH is critical to a person's

well-being. The buffer maintaining the pH of human blood involves carbonic acid (H_2CO_3), bicarbonate ion (HCO_3^-), and carbon dioxide (CO_2). When bicarbonate ions combine with free hydrogen ions and become carbonic acid, hydrogen ions are removed, moderating pH changes. Similarly, as shown in **Figure 11.14**, excess carbonic acid can be converted to carbon dioxide gas and exhaled through the lungs. This prevents too many free hydrogen ions from building up in the blood and dangerously reducing the blood's pH. Likewise, if too much OH^- is introduced into the system, carbonic acid will combine with it to create bicarbonate, lowering the pH. Without this buffer system, the body's pH would fluctuate enough to put survival in jeopardy.

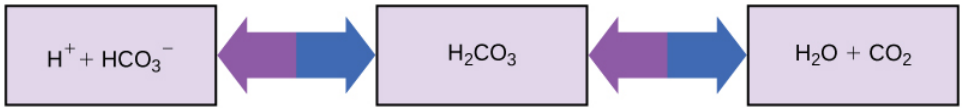


Figure 11.14 This diagram shows the body's buffering of blood pH levels. The blue arrows show the process of raising pH as more CO_2 is produced and exhaled. The purple arrows indicate the reverse process: the lowering of pH as more bicarbonate (HCO_3^-) is generated from the CO_2 being produced by cellular respiration. In short, when more CO_2 is exhaled, blood pH goes up and the blood is more basic, and if CO_2 is converted to bicarbonate, blood pH goes down and the blood is more acidic.

Other examples of buffers are antacids used to combat excess stomach acid. Many of these over-the-counter medications work in the same way as blood buffers, usually with at least one ion capable of absorbing hydrogen and moderating pH, bringing relief to those that suffer "heartburn" after eating. The unique properties of water that contribute to this capacity to balance pH—as well as water's other characteristics—are essential to sustaining life on Earth.

Enzyme Function and Buffers

On the cellular level, cells need to maintain a relatively constant internal environment, because many important cellular functions only take place within a narrow of conditions, such as temperature and pH. Enzymes function is influenced by the pH of its environment and only function optimally within a narrow pH range. Losing function of just one enzyme can mean death for a cell, or tissue, or organ or and entire organism. In addition, there are many reactions occurring in a cell that either produces or consumes Hydrogen ions thereby changing the pH of the cells environment. How do cells maintain pH within the range necessary for enzymes to function properly? The answer is buffers, just as learned earlier about your whole body using buffers to maintain its pH.

12 | TOUR OF THE CELL: NUCLEIC ACIDS AND CELL CYCLE

12.1 | Nucleic Acids and Nucleotides

> "If the results of the present study on the chemical nature of the transforming principle are confirmed, then nucleic acids must be regarded as possessing biological specificity, the nature of which is as yet undetermined."

O.T. Avery, C. MacLeod, and M McCarty, "Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types", *Journal of Experimental Medicine* 1944, Volume 79, p. 155.

In 1944, when Avery, MacLeod and McCarty published the work quoted above, the structure of nucleic acids was still unknown. But their work was one of the more important bits of evidence indicating that nucleic acids contained important biological information. We now know a lot more about the information content, and, thanks to the work of Watson and Crick a few years after this statement was published, we know how the structure of nucleic acids is related to the function of information storage, We know that **nucleic acids** are the most important macromolecules for the continuity of life. They carry the genetic blueprint of a cell, and carry instructions for the functioning of the cell.

DNA and RNA

The two main types of nucleic acids are **deoxyribonucleic acid (DNA)** and **ribonucleic acid (RNA)**. DNA is the genetic material found in all living organisms, ranging from single-celled bacteria to multicellular mammals. It is found in the nucleus of eukaryotes and in the organelles, chloroplasts, and mitochondria. In prokaryotes, the DNA is not enclosed in a membranous envelope.

The entire genetic content of a cell is known as its **genome**, and the study of genomes is genomics. In eukaryotic cells but not in prokaryotes, DNA forms a complex with histone proteins to form **chromatin**, the substance of eukaryotic chromosomes. A chromosome may contain tens of thousands of genes. Many genes contain the information to make protein products; other genes code for RNA products. DNA controls all of the cellular activities by turning the genes "on" or "off."

The other type of nucleic acid, RNA, is mostly involved in protein synthesis. The DNA molecules never leave the nucleus but instead use an intermediary to communicate with the rest of the cell. This intermediary is the **messenger RNA (mRNA)**. Other types of RNA—like rRNA, tRNA, and microRNA—are involved in protein synthesis and its regulation.

DNA and RNA are made up of monomers known as **nucleotides**. The nucleotides combine with each other to form a polynucleotide or **nucleic acid**, DNA or RNA. Each nucleotide is made up of three components: a nitrogenous base, a pentose (five-carbon) sugar, and a phosphate group (**Figure 12.1**). Each nitrogenous base in a nucleotide is attached to a sugar molecule, which is attached to one or more phosphate groups. Some nucleotides such as ATP (Adenosine Triphosphate) are a short term energy source for all cells and is called the "Energy currency of the cell."

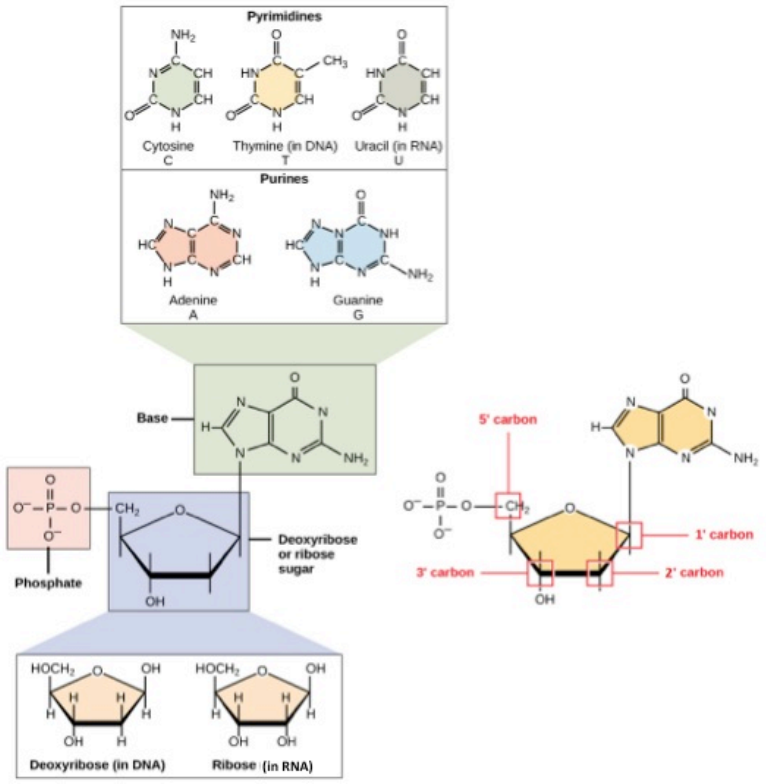


Figure 12.1 A nucleotide is made up of three components: a nitrogenous base, a pentose sugar, and one or more phosphate groups. Carbon residues in the pentose are numbered 1′ through 5′ (the prime distinguishes these residues from those in the base, which are numbered without using a prime notation). The base is attached to the 1′ position of the ribose, and the phosphate is attached to the 5′ position. When a polynucleotide is formed, the 5′ phosphate of the incoming nucleotide attaches to the 3′ hydroxyl group at the end of the growing chain. Two types of pentose are found in nucleotides, deoxyribose (found in DNA) and ribose (found in RNA). Deoxyribose is similar in structure to ribose, but it has an H instead of an OH at the 2′ position. Bases can be divided into two categories: purines and pyrimidines. Purines have a double ring structure, and pyrimidines have a single ring.

The nitrogenous bases, important components of nucleotides, are organic molecules and are so named because they contain carbon and nitrogen. They are bases because they contain an amino group that has the potential of binding an extra hydrogen, and thus, decreases the hydrogen ion concentration in its environment, making it more basic. Each nucleotide in DNA contains one of four possible nitrogenous bases: adenine (A), guanine (G) cytosine (C), and thymine (T).

Adenine and guanine are classified as purines. The primary structure of a purine is two carbon-nitrogen rings. Cytosine, thymine, and uracil are classified as pyrimidines which have a single carbon-nitrogen ring as their primary structure (**Figure 12.1**). Each of these basic carbon-nitrogen rings has different functional groups attached to it. In molecular biology shorthand, the nitrogenous bases are simply known by their symbols A, T, G, C, and U. DNA contains A, T, G, and C whereas RNA contains A, U, G, and C.

The pentose sugar in DNA is deoxyribose, and in RNA, the sugar is ribose (**Figure 12.1**). The difference between the sugars is the presence of the hydroxyl group on the second carbon of the ribose and hydrogen on the second carbon of the deoxyribose. The carbon atoms of the sugar molecule are numbered as 1′, 2′, 3′, 4′, and 5′ (1′ is read as “one prime”). The phosphate residue is attached to the hydroxyl group of the 5′ carbon of one sugar and the hydroxyl group of the 3′ carbon of the sugar of the next nucleotide, which forms a 5′–3′ phosphodiester bond. The phosphodiester bond is not formed by simple condensation reaction like the other linkages connecting monomers in macromolecules: its formation involves the removal of two phosphate groups. A polynucleotide may have thousands of such phosphodiester bonds.

DNA Double-Helix Structure

DNA has a double-helix structure (**Figure 12.2**). The sugar and phosphate lie on the outside of the helix, forming the backbone of the DNA. The nitrogenous bases are stacked in the interior, like the steps of a staircase, in pairs; the pairs are bound to each other by hydrogen bonds. Every base pair in the double helivx is separated from the next base pair by 0.34 nm. The two strands of the helix run in opposite directions, meaning that the 5′ carbon end of one strand will face the 3′

carbon end of its matching strand. (This is referred to as antiparallel orientation and is important to DNA replication and in many nucleic acid interactions.)

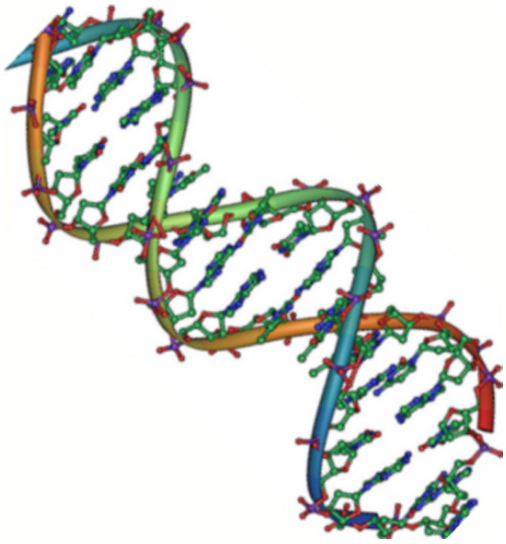


Figure 12.2 Native DNA is an antiparallel double helix. The phosphate backbone (indicated by the curvy lines) is on the outside, and the bases are on the inside. Each base from one strand interacts via hydrogen bonding with a base from the opposing strand. (credit: Jerome Walker/Dennis Myts)

Only certain types of base pairing are allowed. For example, a certain purine can only pair with a certain pyrimidine. This means A can pair with T, and G can pair with C, as shown in **Figure 12.3**. This is known as the base complementary rule. In other words, the DNA strands are complementary to each other. If the sequence of one strand is AATTGGCC, the complementary strand would have the sequence TTAACCGG. During DNA replication, each strand is copied, resulting in a daughter DNA double helix containing one parental DNA strand and a newly synthesized strand.

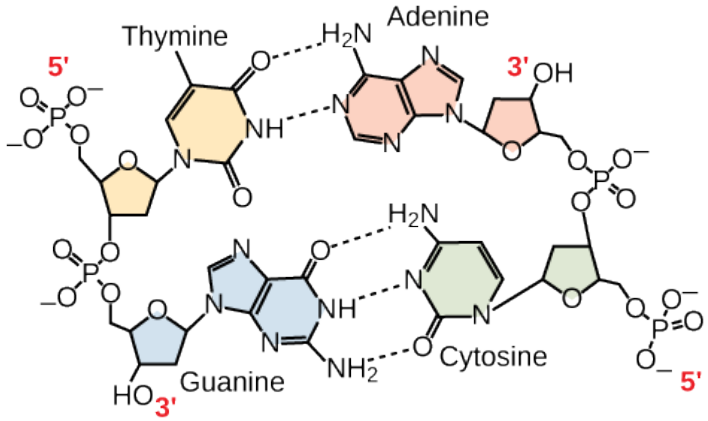


Figure 12.3 In a double stranded DNA molecule, the two strands run antiparallel to one another so that one strand runs 5′ to 3′ and the other 3′ to 5′. The phosphate backbone is located on the outside, and the bases are in the middle. Adenine forms hydrogen bonds (or base pairs) with thymine, and guanine base pairs with cytosine.

RNA

Ribonucleic acid, or RNA, is mainly involved in the process of protein synthesis under the direction of DNA. RNA is usually single-stranded and is made of ribonucleotides that are linked by phosphodiester bonds. A ribonucleotide in the RNA chain contains ribose (the pentose sugar), one of the four nitrogenous bases (A, U, G, and C), and the phosphate group. The function of RNA is to make proteins from the information stored on the DNA.

Features of DNA and RNA

	DNA	RNA
Function	Carries genetic information	Involved in protein synthesis
Location	Remains in the nucleus	Leaves the nucleus
Structure	Double helix	Usually single-stranded
Sugar	Deoxyribose	Ribose
Pyrimidines	Cytosine, thymine	Cytosine, uracil
Purines	Adenine, guanine	Adenine, guanine

Table 12.1

12.2 | Cell Reproduction

By the end of this section, you will be able to:

- Describe the structure of prokaryotic and eukaryotic genomes
- Distinguish between chromosomes, genes, and traits
- Describe the mechanisms of chromosome compaction

Introduction

> "DNA was the first three-dimensional Xerox machine."

Kenneth E. Boulding, in Richard P. Beilock (ed.), *Beasts, Ballads and Bouldingisms: A Selection of Writings by Kenneth E. Boulding*, pg. 160, 1976

Non scientists like Boulding , who was an economist, are fond of using machines as metaphors for biological structures and functions. But DNA is a lot more than just a Xerox machine, as you will learn in this section. In fact, the more we learn about DNA, the more remarkable it seems. For DNA not only replicates itself, it directs the development, life, and demise of every organism on the planet.

A human, as well as every sexually reproducing organism, begins life as a fertilized egg (embryo) or zygote. Trillions of cell divisions subsequently occur in a controlled manner to produce a complex, multicellular human. In other words, that original single cell is the ancestor of every other cell in the body. Once a being is fully grown, cell reproduction is still necessary to repair or regenerate tissues. For example, new blood and skin cells are constantly being produced. All multicellular organisms use cell division for growth and the maintenance and repair of cells and tissues. Cell division is tightly regulated, and the occasional failure of regulation can have life-threatening consequences. Single-celled organisms use cell division as their method of reproduction.

So, the continuity of life from one cell to another has its foundation in the reproduction of cells by way of the cell cycle. The **cell cycle** is an orderly sequence of events that describes the stages of a cell's life from the division of a single parent cell to the production of two new daughter cells. The mechanisms involved in the cell cycle are highly regulated.

Genomic DNA

Before discussing the steps a cell must undertake to replicate, a deeper understanding of the structure and function of a cell's genetic information is necessary. A cell's DNA, packaged as a double-stranded DNA molecule, is called its **genome**. In prokaryotes, the genome is composed of a single, double-stranded DNA molecule in the form of a loop or circle (**Figure 12.4**). The region in the cell containing this genetic material is called a nucleoid. Some prokaryotes also have smaller loops of DNA called **plasmids** that are not essential for normal growth. Bacteria can exchange these plasmids with other bacteria, sometimes receiving beneficial new genes that the recipient can add to their chromosomal DNA. Antibiotic resistance is one trait that often spreads through a bacterial colony through plasmid exchange.

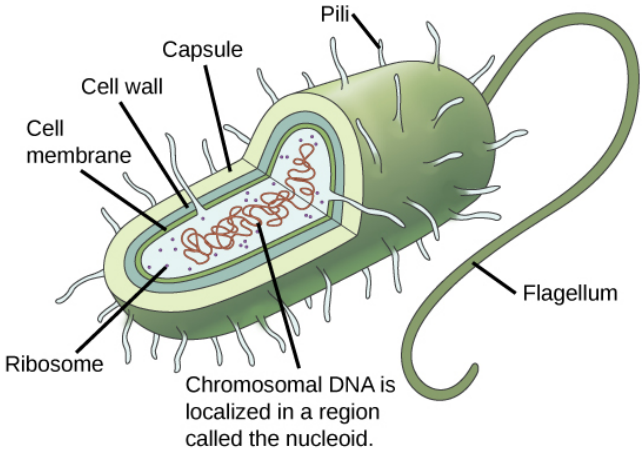


Figure 12.4 Prokaryotes, including bacteria and archaea, have a single, circular chromosome located in a central region called the nucleoid.

In eukaryotes, the genome consists of several double-stranded linear DNA molecules (**Figure 12.5**). Each species of eukaryotes has a characteristic number of chromosomes in the nuclei of its cells. Human body cells have 46 chromosomes, while human **gametes** (sperm or eggs) have 23 chromosomes each. A typical body cell, or somatic cell, contains two matched sets of chromosomes, a configuration known as **diploid**. The letter *n* is used to represent a single set of chromosomes; therefore, a diploid organism is designated 2*n*. Human cells that contain one set of chromosomes are called gametes, or sex cells; these are eggs and sperm, and are designated *1n*, or **haploid**.

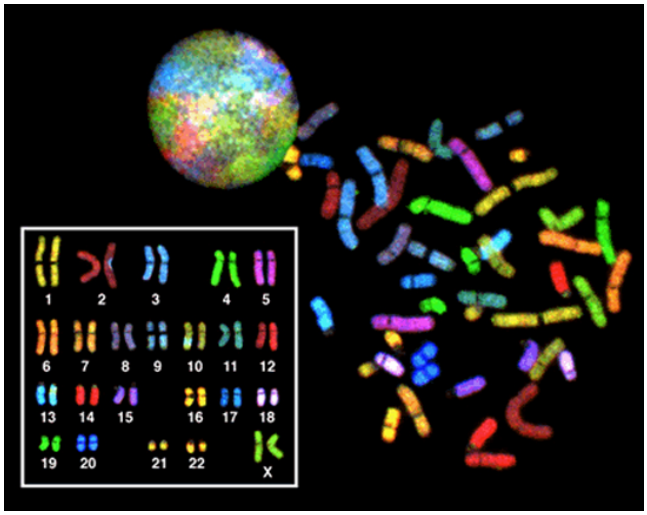


Figure 12.5 There are 23 pairs of homologous chromosomes in a female human somatic cell. The condensed chromosomes are viewed within the nucleus (top), removed from a cell in mitosis and spread out on a slide (right), and artificially arranged according to length (left); an arrangement like this is called a karyotype. In this image, the chromosomes were exposed to fluorescent stains for differentiation of the different chromosomes. A method of staining called "chromosome painting" employs fluorescent dyes that highlight chromosomes in different colors. (credit: National Human Genome Project/NIH)

Matched pairs of chromosomes in a diploid organism are called **homologous** ("same knowledge") **chromosomes**. Homologous chromosomes are the same length and have specific nucleotide segments called **genes** in exactly the same location, or **locus**. Genes, the functional units of chromosomes, determine specific characteristics by coding for specific proteins. Traits are the variations of those characteristics. For example, hair color is a characteristic with traits that are blonde, brown, or black.

Each copy of a homologous pair of chromosomes originates from a different parent; therefore, the genes themselves are not identical. The variation of individuals within a species is due to the specific combination of the genes inherited from both parents. Even a slightly altered sequence of nucleotides within a gene can result in an alternative trait. For example, there are three possible gene sequences on the human chromosome that code for blood type: sequence A, sequence B, and sequence O. Because all diploid human cells have two copies of the chromosome that determines blood type, the blood type

(the trait) is determined by which two versions of the marker gene are inherited. It is possible to have two copies of the same gene sequence on both homologous chromosomes, with one on each (for example, AA, BB, or OO), or two different sequences, such as AB. These different versions of genes are called **alleles**.

Minor variations of traits, such as blood type, eye color, and handedness, contribute to the natural variation found within a species. However, if the entire DNA sequence from any pair of human homologous chromosomes is compared, the difference is less than one percent. The sex chromosomes, X and Y, are the single exception to the rule of homologous chromosome uniformity: Other than a small amount of homology that is necessary to accurately produce gametes, the genes found on the X and Y chromosomes are different.

Eukaryotic Chromosomal Structure and Compaction

If the DNA from all 46 chromosomes in a human cell nucleus was laid out end to end, it would measure approximately two meters; however, its diameter would be only 2 nm. Considering that the size of a typical human cell is about 10 µm (100,000 cells lined up to equal one meter), DNA must be tightly packaged to fit in the cell's nucleus. At the same time, it must also be readily accessible for the genes to be expressed. During some stages of the cell cycle, the long strands of DNA are condensed into compact chromosomes. There are a number of ways that chromosomes are compacted.

In the first level of compaction, short stretches of the DNA double helix wrap around a core of eight histone proteins at regular intervals along the entire length of the chromosome (**Figure 12.6**). The DNA-histone complex is called chromatin. The beadlike, histone DNA complex is called a nucleosome, and DNA connecting the nucleosomes is called linker DNA. A DNA molecule in this form is about seven times shorter than the double helix without the histones, and the beads are about 10 nm in diameter, in contrast with the 2-nm diameter of a DNA double helix. The next level of compaction occurs as the nucleosomes and the linker DNA between them are coiled into a 30-nm chromatin fiber. This coiling further shortens the chromosome so that it is now about 50 times shorter than the extended form. In the third level of packing, a variety of fibrous proteins is used to pack the chromatin. These fibrous proteins also ensure that each chromosome in a non-dividing cell occupies a particular area of the nucleus that does not overlap with that of any other chromosome (see the top image in **Figure 12.5**).

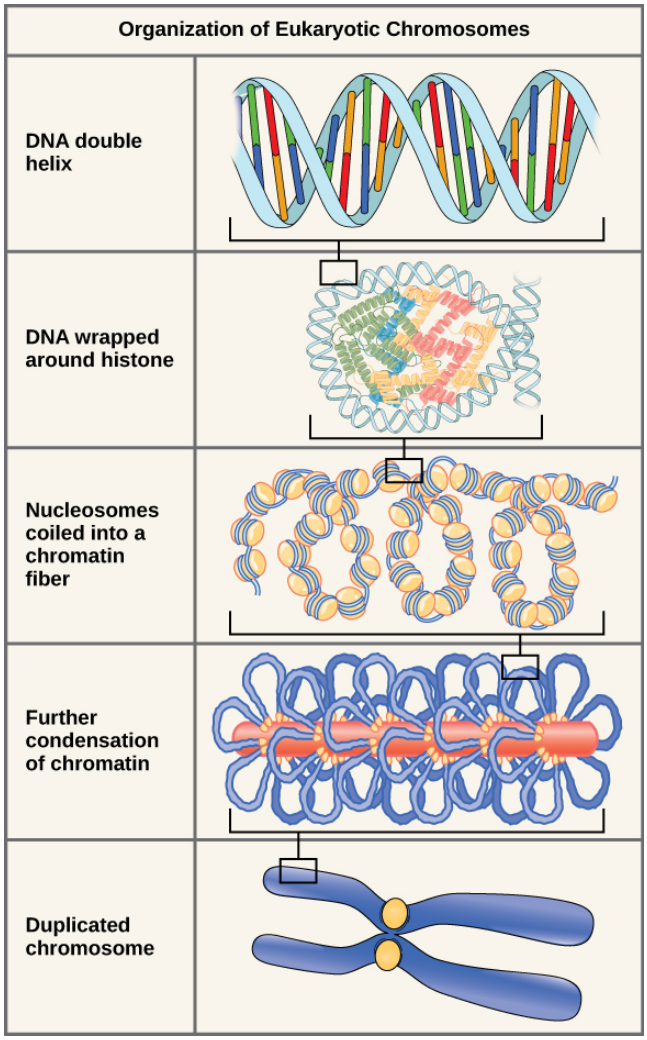


Figure 12.6 Double-stranded DNA wraps around histone proteins to form nucleosomes that have the appearance of "beads on a string." The nucleosomes are coiled into a 30-nm chromatin fiber. When a cell undergoes mitosis, the chromosomes condense even further.

DNA replicates in the S phase of interphase. After replication, the chromosomes are composed of two linked sister **chromatids**. When fully compact, the pairs of identically packed chromosomes are bound to each other by cohesin proteins. The connection between the sister chromatids is closest in a region called the **centromere**. The conjoined sister chromatids, with a diameter of about 1 µm, are visible under a light microscope. The centromeric region is highly condensed and thus will appear as a constricted area.

12.3 | DNA Replication

By the end of this section, you will be able to:

- Explain the process of DNA replication
- Explain the importance of telomerase to DNA replication
- Describe mechanisms of DNA repair

When a cell divides, it is important that each daughter cell receives an identical copy of the DNA. This is accomplished by the process of DNA replication. The replication of DNA occurs during the synthesis phase, or S phase, of the cell cycle, before the cell enters mitosis or meiosis.

The elucidation of the structure of the double helix provided a hint as to how DNA is copied. Recall that adenine nucleotides

pair with thymine nucleotides, and cytosine with guanine. This means that the two strands are complementary to each other. For example, a strand of DNA with a nucleotide sequence of AGTCATGA will have a complementary strand with the sequence TCAGTACT (**Figure 12.7**).

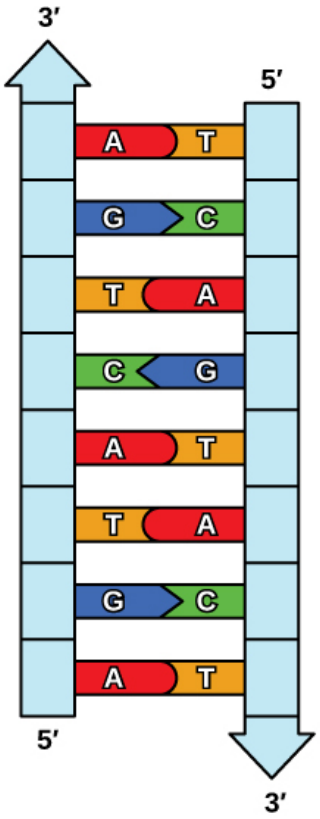


Figure 12.7 The two strands of DNA are complementary, meaning the sequence of bases in one strand can be used to create the correct sequence of bases in the other strand.

Because of the complementarity of the two strands, having one strand means that it is possible to recreate the other strand. This model for replication suggests that the two strands of the double helix separate during replication, and each strand serves as a template from which the new complementary strand is copied (**Figure 12.8**).

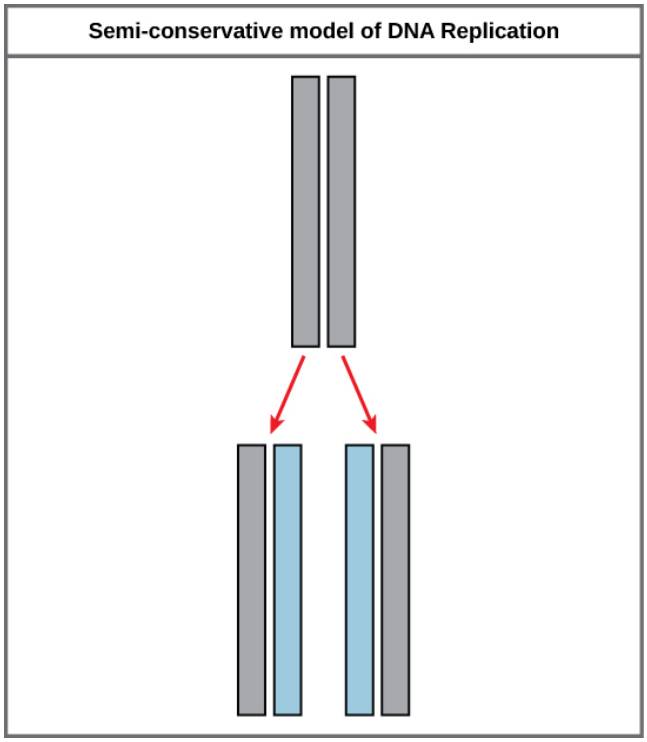


Figure 12.8 The semiconservative model of DNA replication is shown. Gray indicates the original DNA strands, and blue indicates newly synthesized DNA.

During DNA replication, each of the two strands that make up the double helix serves as a template from which new strands are copied. The new strand will be complementary to the parental or “old” strand. Each new double strand consists of one parental strand and one new daughter strand. This is known as **semiconservative replication**. When two DNA copies are formed, they have an identical sequence of nucleotide bases and are divided equally into two daughter cells.

DNA Replication in Eukaryotes

Because eukaryotic genomes are very complex, DNA replication is a very complicated process that involves several enzymes and other proteins. It occurs in three main stages: initiation, elongation, and termination.

Recall that eukaryotic DNA is bound to proteins known as histones to form structures called nucleosomes. During initiation, the DNA is made accessible to the proteins and enzymes involved in the replication process. How does the replication machinery know where on the DNA double helix to begin? It turns out that there are specific nucleotide sequences called origins of replication at which replication begins. Certain proteins bind to the origin of replication while an enzyme called **helicase** unwinds and opens up the DNA helix. As the DNA opens up, Y-shaped structures called **replication forks** are formed (**Figure 12.9**). Two replication forks are formed at the origin of replication, and these get extended in both directions as replication proceeds. There are multiple origins of replication on the eukaryotic chromosome, such that replication can occur simultaneously from several places in the genome.

During elongation, an enzyme called **DNA polymerase** adds DNA nucleotides to the 3' end of the template. Because DNA polymerase can only add new nucleotides at the end of a backbone, a **primer** sequence, which provides this starting point, is added with complementary RNA nucleotides. This primer is removed later, and the nucleotides are replaced with DNA nucleotides. One strand, which is complementary to the parental DNA strand, is synthesized continuously toward the replication fork so the polymerase can add nucleotides in this direction. This continuously synthesized strand is known as the **leading strand**. Because DNA polymerase can only synthesize DNA in a 5' to 3' direction, the other new strand is put together in short pieces called **Okazaki fragments**. The Okazaki fragments each require a primer made of RNA to start the synthesis. The strand with the Okazaki fragments is known as the **lagging strand**. As synthesis proceeds, an enzyme removes the RNA primer, which is then replaced with DNA nucleotides, and the gaps between fragments are sealed by an enzyme called **DNA ligase**.

The process of DNA replication can be summarized as follows:

1. DNA unwinds at the origin of replication.
2. New bases are added to the complementary parental strands. One new strand is made continuously, while the other

strand is made in pieces.

3. Primers are removed, new DNA nucleotides are put in place of the primers and the backbone is sealed by DNA ligase.

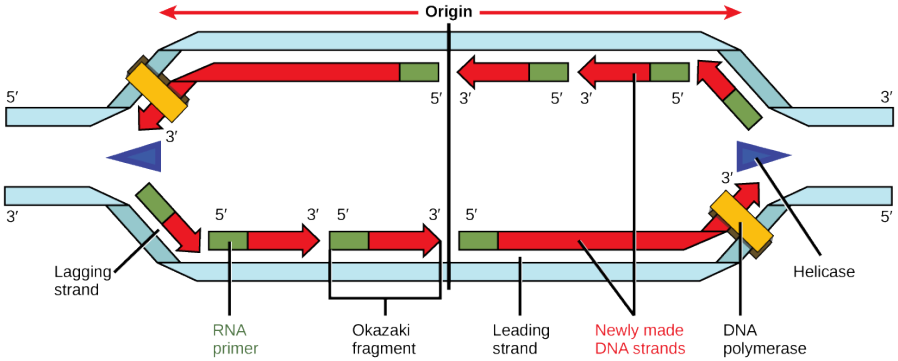


Figure 12.9 A replication fork is formed by the opening of the origin of replication, and helicase separates the DNA strands. An RNA primer is synthesized, and is elongated by the DNA polymerase. On the leading strand, DNA is synthesized continuously, whereas on the lagging strand, DNA is synthesized in short stretches. The DNA fragments are joined by DNA ligase (not shown).

Telomere Replication

Because eukaryotic chromosomes are linear, DNA replication comes to the end of a line in eukaryotic chromosomes. As you have learned, the DNA polymerase enzyme can add nucleotides in only one direction. In the leading strand, synthesis continues until the end of the chromosome is reached; however, on the lagging strand there is no place for a primer to be made for the DNA fragment to be copied at the end of the chromosome. This presents a problem for the cell because the ends remain unpaired, and over time these ends get progressively shorter as cells continue to divide. The ends of the linear chromosomes are known as **telomeres**, which have repetitive sequences that do not code for a particular gene. As a consequence, it is telomeres that are shortened with each round of DNA replication instead of genes. For example, in humans, a six base-pair sequence, TTAGGG, is repeated 100 to 1000 times. The discovery of the enzyme **telomerase** (**Figure 12.10**) helped in the understanding of how chromosome ends are maintained. The telomerase attaches to the end of the chromosome, and complementary bases to the RNA template are added on the end of the DNA strand. Once the lagging strand template is sufficiently elongated, DNA polymerase can now add nucleotides that are complementary to the ends of the chromosomes. Thus, the ends of the chromosomes are replicated.

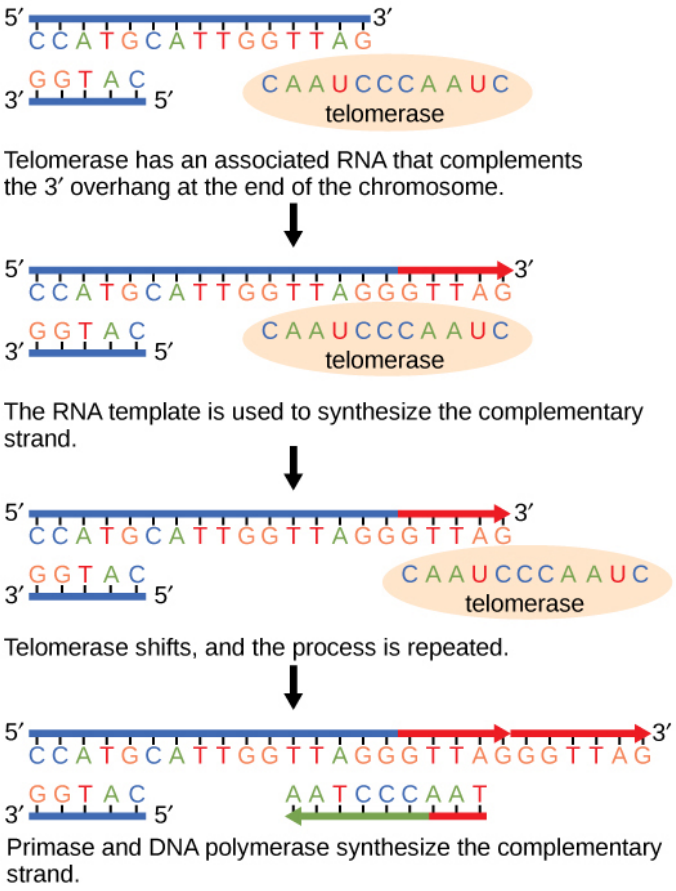


Figure 12.10 The ends of linear chromosomes are maintained by the action of the telomerase enzyme.

Telomerase is typically found to be active in germ cells, adult stem cells, and some cancer cells. For her discovery of telomerase and its action, Elizabeth Blackburn (**Figure 12.11**) received the Nobel Prize for Medicine and Physiology in 2009.

Figure 12.11 Elizabeth Blackburn, 2009 Nobel Laureate, was the scientist who discovered how telomerase works. (credit: U.S. Embassy, Stockholm, Sweden)

Telomerase is not active in adult somatic cells. Adult somatic cells that undergo cell division continue to have their telomeres shortened. This essentially means that telomere shortening is associated with aging. In 2010, scientists found that telomerase can reverse some age-related conditions in mice, and this may have potential in regenerative medicine.[1]

Telomerase-deficient mice were used in these studies; these mice have tissue atrophy, stem-cell depletion, organ system failure, and impaired tissue injury responses. Telomerase reactivation in these mice caused extension of telomeres, reduced DNA damage, reversed neurodegeneration, and improved functioning of the testes, spleen, and intestines. Thus, telomere reactivation may have potential for treating age-related diseases in humans.

DNA Repair

DNA polymerase can make mistakes while adding nucleotides. It edits the DNA by proofreading every newly added base. Incorrect bases are removed and replaced by the correct base, and then polymerization continues (**Figure 12.12a**). Most mistakes are corrected during replication, although when this does not happen, the mismatch repair mechanism is employed. Mismatch repair enzymes recognize the wrongly incorporated base and excise it from the DNA, replacing it with the correct base (**Figure 12.12b**). In yet another type of repair, nucleotide excision repair, the DNA double strand is unwound and separated, the incorrect bases are removed along with a few bases on the 5' and 3' end, and these are replaced by copying the template with the help of DNA polymerase (**Figure 12.12c**). Nucleotide excision repair is particularly important in correcting thymine dimers, which are primarily caused by ultraviolet light. In a thymine dimer, two thymine nucleotides adjacent to each other on one strand are covalently bonded to each other rather than their complementary bases. If the dimer is not removed and repaired it will lead to a mutation. Individuals with flaws in their nucleotide excision repair genes show extreme sensitivity to sunlight and develop skin cancers early in life.

1. Mariella Jaskelioff, et al., "Telomerase reactivation reverses tissue degeneration in aged telomerase-deficient mice," *Nature*, 469 (2011):102–7.

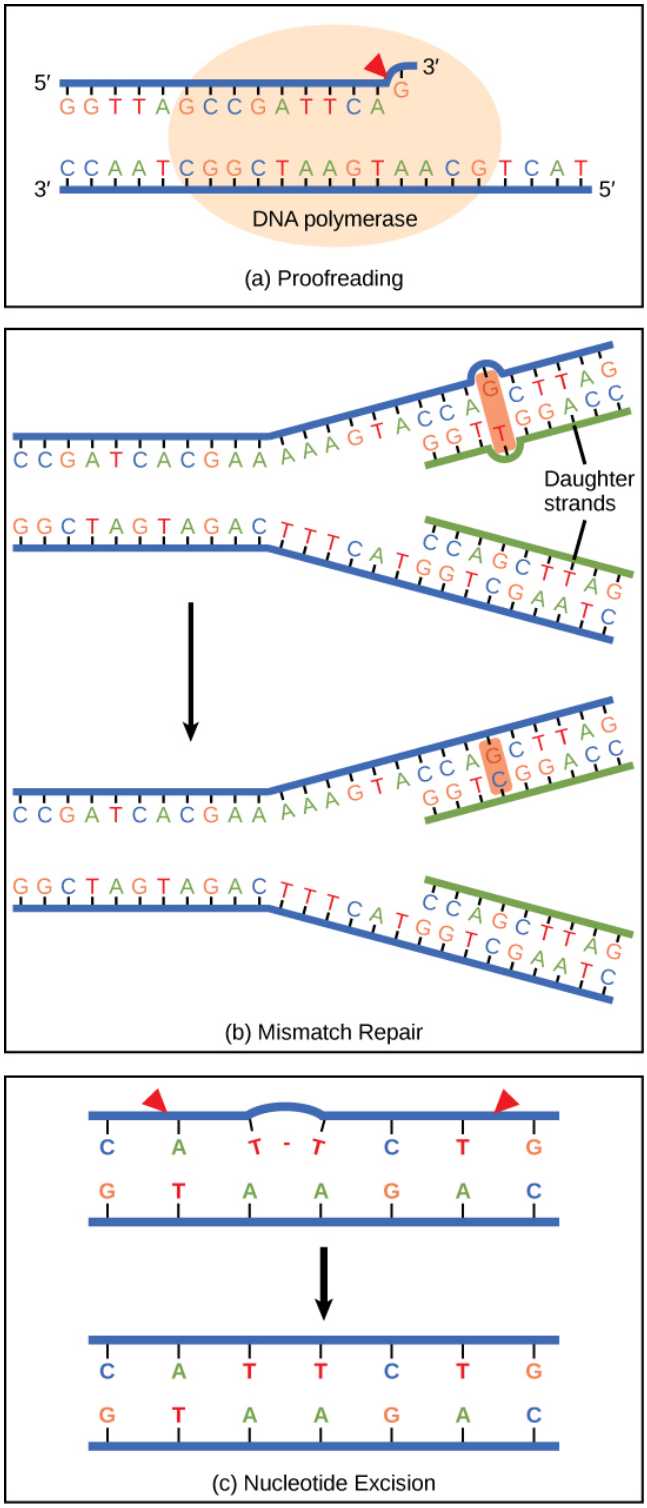


Figure 12.12 Proofreading by DNA polymerase (a) corrects errors during replication. In mismatch repair (b), the incorrectly added base is detected after replication. The mismatch repair proteins detect this base and remove it from the newly synthesized strand by nuclease action. The gap is now filled with the correctly paired base. Nucleotide excision (c) repairs thymine dimers. When exposed to UV, thymines lying adjacent to each other can form thymine dimers. In normal cells, they are excised and replaced.

Most mistakes are corrected; if they are not, they may result in a **mutation**—defined as a permanent change in the DNA sequence. Mutations in repair genes may lead to serious consequences like cancer.

12.4 | Prokaryotic Cell Division

Introduction

> “...(T)he extreme rapidity with which generation succeeds generation amongst bacteria offers to the forces of variation and natural selection a field for their operation wholly unparalleled amongst higher forms of life.”

Frederick W. Andrewes, "The Evolution of the Streptococci", *The Lancet*, 2:1415, 1906

Prokaryotes such as bacteria propagate by **binary fission**, and produces two identical daughter cells. This is a simpler process than cell division in eukaryotes, partly because there is no need for separate nuclear and cellular divisions (mitosis and cytokinesis, respectively). As Andrewes noted, the rapidity of cell reproduction in prokaryotes opens the door for rapid evolutionary change in these organisms.

To achieve the outcome of identical daughter cells, some steps are essential. The genomic DNA must be replicated and then allocated into the daughter cells; the cytoplasmic contents must also be divided to give both new cells the machinery to sustain life. In bacterial cells, the genome consists of a single, circular DNA chromosome; therefore, the process of cell division is simplified. Mitosis is unnecessary because there is no nucleus or multiple chromosomes. This type of cell division is called binary fission.

Binary Fission

The cell division process of prokaryotes, called **binary fission**, is a less complicated and much quicker process than cell division in eukaryotes. Because of the speed of bacterial cell division, populations of bacteria can grow very rapidly. The single, circular DNA chromosome of bacteria is not enclosed in a nucleus, but instead occupies a specific location, the nucleoid, within the cell. As in eukaryotes, the DNA of the nucleoid is associated with proteins that aid in packaging the molecule into a compact size. The packing proteins of bacteria are, however, related to some of the proteins involved in the chromosome compaction of eukaryotes.

The starting point of replication, the origin, is close to the binding site of the chromosome to the plasma membrane (**Figure 12.13**). Replication of the DNA is bidirectional—moving away from the origin on both strands of the DNA loop simultaneously. As the new double strands are formed, each origin point moves away from the cell-wall attachment toward opposite ends of the cell. As the cell elongates, the growing membrane aids in the transport of the chromosomes. After the chromosomes have cleared the midpoint of the elongated cell, cytoplasmic separation begins. A septum is formed between the nucleoids from the periphery toward the center of the cell. When the new cell walls are in place, the daughter cells separate.

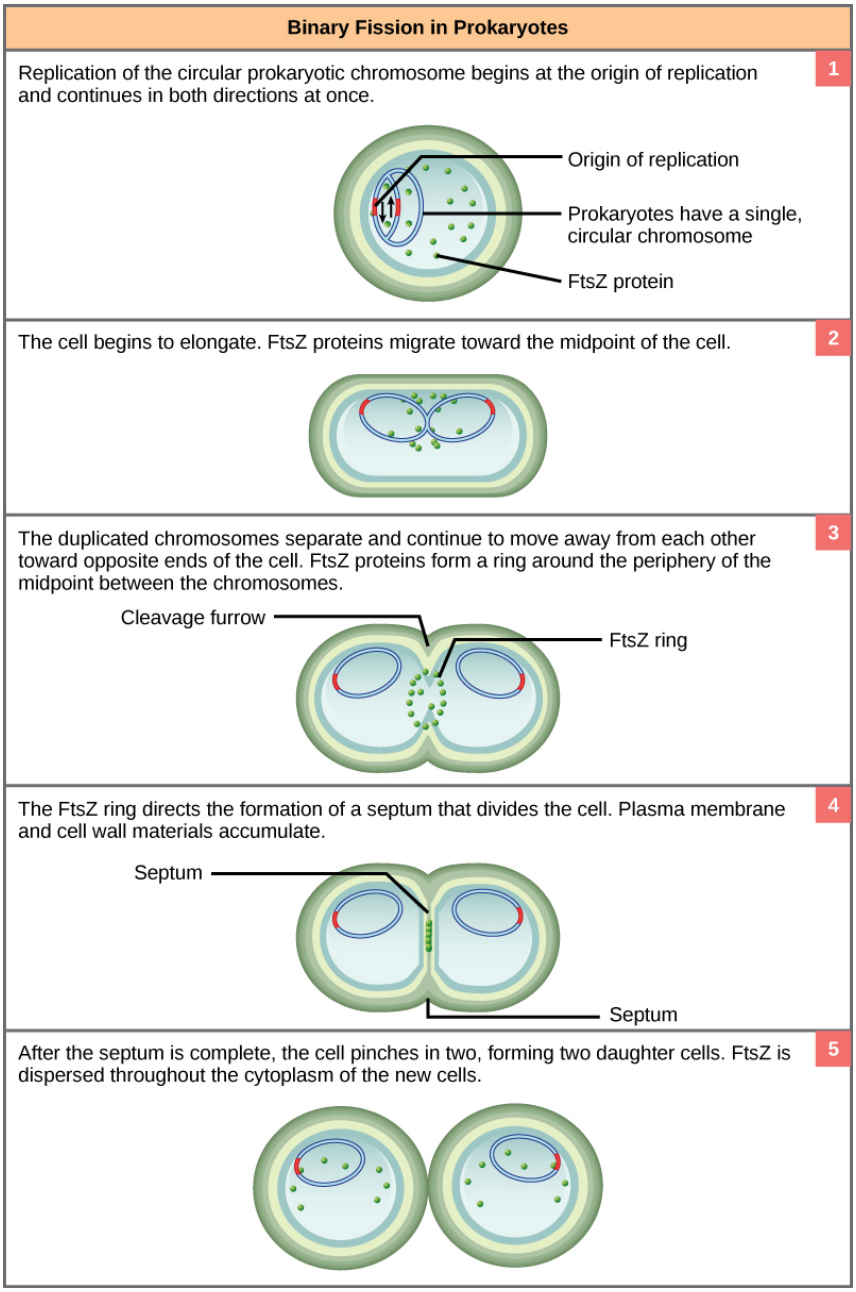


Figure 12.13 The binary fission of a bacterium is outlined in five steps. (credit: modification of work by "Mcstrother"/Wikimedia Commons)

evolution IN ACTION

Mitotic Spindle Apparatus

The precise timing and formation of the mitotic spindle is critical to the success of eukaryotic cell division. Prokaryotic cells, on the other hand, do not undergo mitosis and therefore have no need for a mitotic spindle. However, the FtsZ protein that plays such a vital role in prokaryotic cytokinesis is structurally and functionally very similar to tubulin, the building block of the microtubules that make up the mitotic spindle fibers that are necessary for eukaryotes. The formation of a ring composed of repeating units of a protein called **FtsZ** directs the partition between the nucleoids in prokaryotes. Formation of the FtsZ ring triggers the accumulation of other proteins that work together to recruit new membrane and cell-wall materials to the site. FtsZ proteins can form filaments, rings, and other three-dimensional structures resembling the way tubulin forms microtubules, centrioles, and various cytoskeleton components. In addition, both FtsZ and tubulin employ the same energy source, GTP (guanosine triphosphate), to rapidly assemble and disassemble complex structures.

FtsZ and tubulin are an example of homology, structures derived from the same evolutionary origins. In this example, FtsZ is presumed to be similar to the ancestor protein to both the modern FtsZ and tubulin. While both proteins are found in extant organisms, tubulin function has evolved and diversified tremendously since the evolution from its FtsZ-like prokaryotic origin. A survey of cell-division machinery in present-day unicellular eukaryotes reveals crucial intermediary steps to the complex mitotic machinery of multicellular eukaryotes (Table 12.2).

Mitotic Spindle Evolution

	Structure of genetic material	Division of nuclear material	Separation of daughter cells
Prokaryotes	There is no nucleus. The single, circular chromosome exists in a region of cytoplasm called the nucleoid.	Occurs through binary fission. As the chromosome is replicated, the two copies move to opposite ends of the cell by an unknown mechanism.	FtsZ proteins assemble into a ring that pinches the cell in two.
Some protists	Linear chromosomes exist in the nucleus.	Chromosomes attach to the nuclear envelope, which remains intact. The mitotic spindle passes through the envelope and elongates the cell. No centrioles exist.	Microfilaments form a cleavage furrow that pinches the cell in two.
Other protists	Linear chromosomes exist in the nucleus.	A mitotic spindle forms from the centrioles and passes through the nuclear membrane, which remains intact. Chromosomes attach to the mitotic spindle. The mitotic spindle separates the chromosomes and elongates the cell.	Microfilaments form a cleavage furrow that pinches the cell in two.

Table 12.2 The mitotic spindle fibers of eukaryotes are composed of microtubules. Microtubules are polymers of the protein tubulin. The FtsZ protein active in prokaryote cell division is very similar to tubulin in the structures it can form and its energy source. Single-celled eukaryotes (such as yeast) display possible intermediary steps between FtsZ activity during binary fission in prokaryotes and the mitotic spindle in multicellular eukaryotes, during which the nucleus breaks down and is reformed.

Mitotic Spindle Evolution

	Structure of genetic material	Division of nuclear material	Separation of daughter cells
Animal cells	Linear chromosomes exist in the nucleus.	A mitotic spindle forms from the centrioles. The nuclear envelope dissolves. Chromosomes attach to the mitotic spindle, which separates them and elongates the cell.	Microfilaments form a cleavage furrow that pinches the cell in two.

Table 12.2 The mitotic spindle fibers of eukaryotes are composed of microtubules. Microtubules are polymers of the protein tubulin. The FtsZ protein active in prokaryote cell division is very similar to tubulin in the structures it can form and its energy source. Single-celled eukaryotes (such as yeast) display possible intermediary steps between FtsZ activity during binary fission in prokaryotes and the mitotic spindle in multicellular eukaryotes, during which the nucleus breaks down and is reformed.

12.5 | Eukaryotic Cell Cycle

Introduction

> “These facts show that mitosis is due to the coordinate play of an extremely complex system of forces which are as yet scarcely comprehended.”

Edmund B. Wilson, *The Cell in Development and Inheritance*, pg. 86, 1896

Wilson was correct in regarding mitosis and cell division as being "extremely complex". But in the century or so that has passed since he wrote those words, we have learned quite a lot about the nature of that complexity. The process has even been given a name: the cell cycle. The **cell cycle** is an ordered series of events involving cell growth and cell division that produces two new daughter cells. Cells on the path to cell division proceed through a series of precisely timed and carefully regulated stages of growth, DNA replication, and division that produce two genetically identical cells. The cell cycle has two major phases: interphase and the mitotic phase (**Figure 12.14**). During **interphase**, the cell grows and DNA is replicated. During the **mitotic phase**, the replicated DNA and cytoplasmic contents are separated and the cell divides.

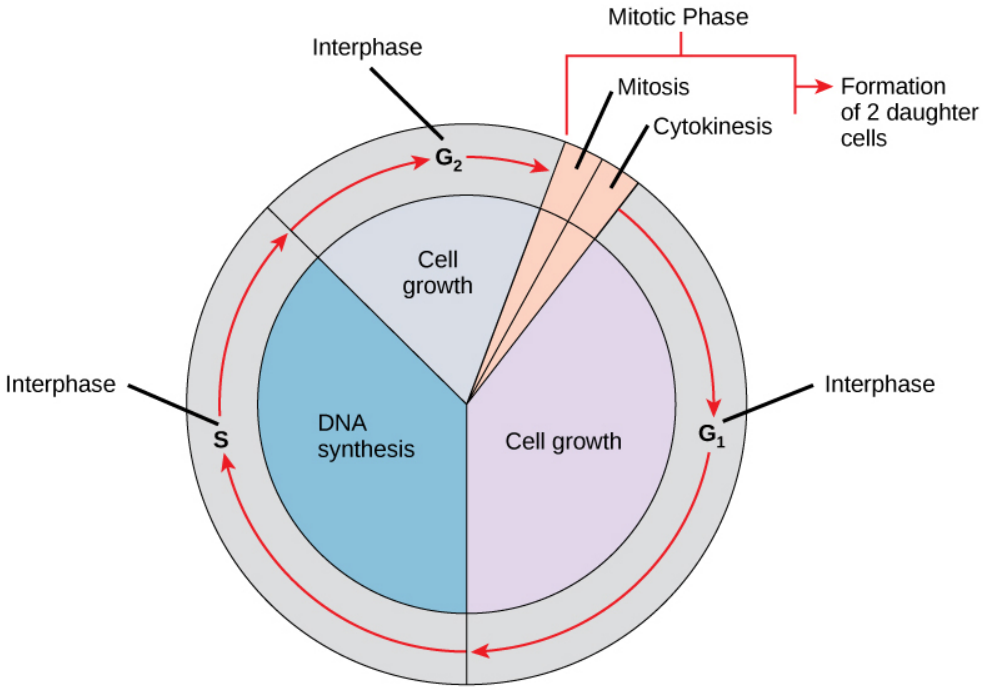


Figure 12.14 A cell moves through a series of phases in an orderly manner. During interphase, G_1 involves cell growth and protein synthesis, the S phase involves DNA replication and the replication of the centrosome, and G_2 involves further growth and protein synthesis. The mitotic phase follows interphase. Mitosis is nuclear division during which duplicated chromosomes are segregated and distributed into daughter nuclei. Usually the cell will divide after mitosis in a process called cytokinesis in which the cytoplasm is divided and two daughter cells are formed.

Interphase

During interphase, the cell undergoes normal processes while also preparing for cell division. For a cell to move from interphase to the mitotic phase, many internal and external conditions must be met. The three stages of interphase are called G_1, S, and G_2.

G_1 Phase

The first stage of interphase is called the **G_1 phase**, or first gap, because little change is visible under the microscope. However, during the G_1 stage, the cell is quite active at the biochemical level. The cell is accumulating the building blocks of chromosomal DNA and the associated proteins, as well as accumulating enough energy reserves to complete the task of replicating each chromosome in the nucleus.

S Phase

Throughout interphase, nuclear DNA remains in a semi-condensed chromatin configuration. In the **S phase** (synthesis phase), DNA replication results in the formation of two identical copies of each chromosome—sister chromatids—that are firmly attached at the centromere region. At this stage, each chromosome is made of two sister chromatids and is a duplicated chromosome. The centrosome is duplicated during the S phase. The two centrosomes will give rise to the **mitotic spindle**, the apparatus that orchestrates the movement of chromosomes during mitosis. The centrosome consists of a pair of rod-like **centrioles** at right angles to each other. Centrioles help organize cell division. Centrioles are not present in the centrosomes of many eukaryotic species, such as plants and most fungi.

G_2 Phase

In the **G_2 phase**, or second gap, the cell replenishes its energy stores and synthesizes the proteins necessary for chromosome manipulation. Some cell organelles are duplicated, and the cytoskeleton is partially dismantled to provide building blocks for the mitotic spindle. There may be additional cell growth during G_2. The final preparations for the mitotic phase must be completed before the cell is able to enter the first stage of mitosis.

The Mitotic Phase

To make two daughter cells, the contents of the nucleus and the cytoplasm must be divided. The mitotic phase is a multistep process during which the duplicated chromosomes are aligned, separated, and moved to opposite poles of the cell, and then

the cell is divided into two new identical daughter cells. The first portion of the mitotic phase, **mitosis**, is composed of five stages, which accomplish nuclear division. The second portion of the mitotic phase, called cytokinesis, is the physical separation of the cytoplasmic components into two daughter cells.

Mitosis

Mitosis is divided into a series of phases—prophase, metaphase, anaphase, and telophase—that result in the division of the cell nucleus (**Figure 12.15**).

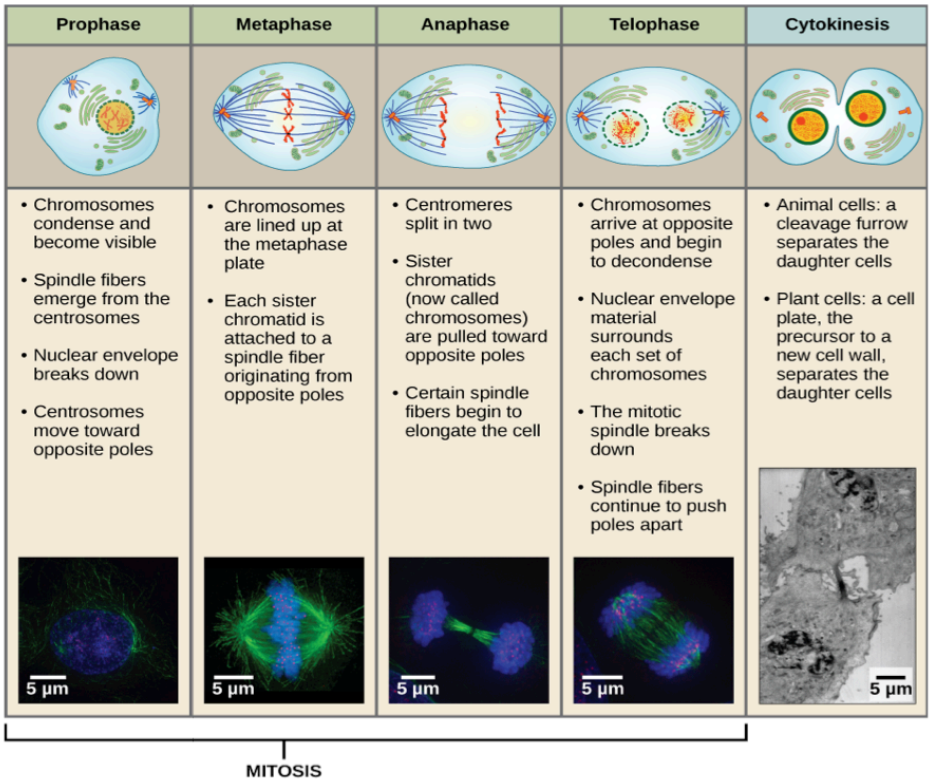


Figure 12.15 Animal cell mitosis is divided into four stages—prophase, metaphase, anaphase, and telophase—visualized here by light microscopy with fluorescence. Mitosis is usually accompanied by cytokinesis, shown here by a transmission electron microscope. (credit "diagrams": modification of work by Mariana Ruiz Villareal; credit "mitosis micrographs": modification of work by Roy van Heesbeen; credit "cytokinesis micrograph": modification of work by the Wadsworth Center, NY State Department of Health; donated to the Wikimedia foundation; scale-bar data from Matt Russell)

During **prophase**, the “first phase,” several events must occur to provide access to the chromosomes in the nucleus. The nuclear envelope starts to break into small vesicles, and the Golgi apparatus and endoplasmic reticulum fragment and disperse to the periphery of the cell. The nucleolus disappears. The centrosomes begin to move to opposite poles of the cell. The microtubules that form the basis of the mitotic spindle extend between the centrosomes, pushing them farther apart as the microtubule fibers lengthen. The sister chromatids begin to coil more tightly and become visible under a light microscope. Also, each sister chromatid attaches to spindle microtubules at the centromere via a protein complex called the kinetochore.

During **metaphase**, all of the chromosomes are aligned in a plane called the **metaphase plate**, or the equatorial plane, midway between the two poles of the cell. The sister chromatids are still tightly attached to each other. At this time, the chromosomes are maximally condensed.

During **anaphase**, the sister chromatids at the equatorial plane are split apart at the centromere. Each chromatid, now called a chromosome, is pulled rapidly toward the centrosome to which its microtubule was attached. The cell becomes visibly elongated as the non-kinetochore microtubules slide against each other at the metaphase plate where they overlap.

During **telophase**, all of the events that set up the duplicated chromosomes for mitosis during the first three phases are reversed. The chromosomes reach the opposite poles and begin to decondense (unravel). The mitotic spindles are broken down into monomers that will be used to assemble cytoskeleton components for each daughter cell. Nuclear envelopes form around chromosomes.

Cytokinesis

Cytokinesis is the second part of the mitotic phase during which cell division is completed by the physical separation of the cytoplasmic components into two daughter cells. Although the stages of mitosis are similar for most eukaryotes, the process of cytokinesis is quite different for eukaryotes that have cell walls, such as plant cells.

In cells such as animal cells that lack cell walls, cytokinesis begins following the onset of anaphase. A contractile ring composed of actin filaments forms just inside the plasma membrane at the former metaphase plate. The actin filaments pull the equator of the cell inward, forming a fissure. This fissure, or "crack," is called the **cleavage furrow**. The furrow deepens as the actin ring contracts, and eventually the membrane and cell are cleaved in two (Figure 12.16).

In plant cells, a cleavage furrow is not possible because of the rigid cell walls surrounding the plasma membrane. A new cell wall must form between the daughter cells. During interphase, the Golgi apparatus accumulates enzymes, structural proteins, and glucose molecules prior to breaking up into vesicles and dispersing throughout the dividing cell. During telophase, these Golgi vesicles move on microtubules to collect at the metaphase plate. There, the vesicles fuse from the center toward the cell walls; this structure is called a **cell plate**. As more vesicles fuse, the cell plate enlarges until it merges with the cell wall at the periphery of the cell. Enzymes use the glucose that has accumulated between the membrane layers to build a new cell wall of cellulose. The Golgi membranes become the plasma membrane on either side of the new cell wall (Figure 12.16).

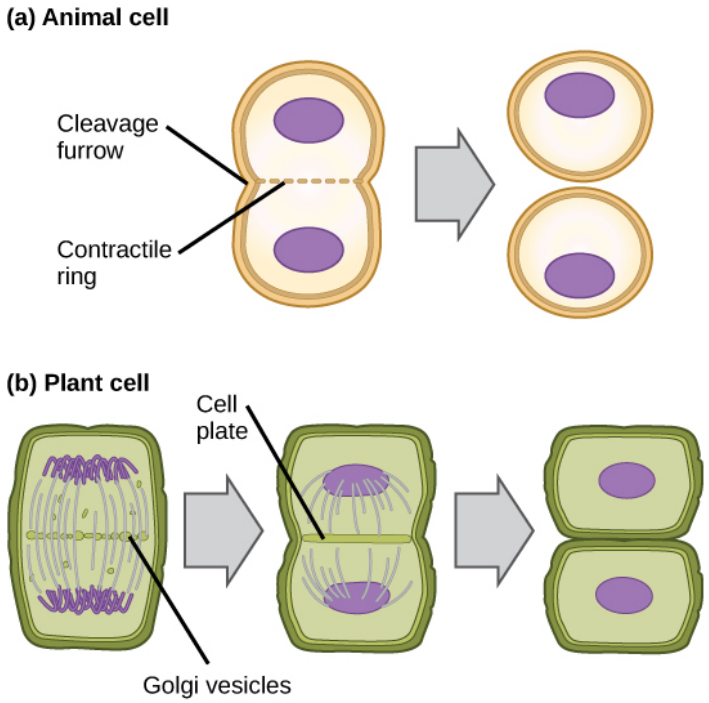


Figure 12.16 In part (a), a cleavage furrow forms at the former metaphase plate in the animal cell. The plasma membrane is drawn in by a ring of actin fibers contracting just inside the membrane. The cleavage furrow deepens until the cells are pinched in two. In part (b), Golgi vesicles coalesce at the former metaphase plate in a plant cell. The vesicles fuse and form the cell plate. The cell plate grows from the center toward the cell walls. New cell walls are made from the vesicle contents.

G_0 Phase

Not all cells adhere to the classic cell-cycle pattern in which a newly formed daughter cell immediately enters interphase, closely followed by the mitotic phase. Cells in the **G_0 phase** are not actively preparing to divide. The cell is in a quiescent (inactive) stage, having exited the cell cycle. Some cells enter G_0 temporarily until an external signal triggers the onset of G_1. Other cells that never or rarely divide, such as mature cardiac muscle and nerve cells, remain in G_0 permanently (Figure 12.17).

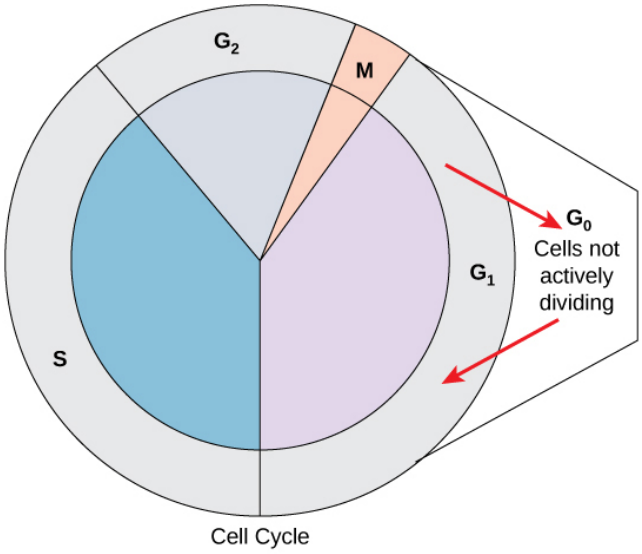


Figure 12.17 Cells that are not actively preparing to divide enter an alternate phase called G_0. In some cases, this is a temporary condition until triggered to enter G_1. In other cases, the cell will remain in G_0 permanently.

Control of the Cell Cycle

The length of the cell cycle is highly variable even within the cells of an individual organism. In humans, the frequency of cell turnover ranges from a few hours in early embryonic development to an average of two to five days for epithelial cells, or to an entire human lifetime spent in G_0 by specialized cells such as cortical neurons or cardiac muscle cells. There is also variation in the time that a cell spends in each phase of the cell cycle. When fast-dividing mammalian cells are grown in culture (outside the body under optimal growing conditions), the length of the cycle is approximately 24 hours. In rapidly dividing human cells with a 24-hour cell cycle, the G_1 phase lasts approximately 11 hours. The timing of events in the cell cycle is controlled by mechanisms that are both internal and external to the cell.

Regulation at Internal Checkpoints

It is essential that daughter cells be exact duplicates of the parent cell. Mistakes in the duplication or distribution of the chromosomes lead to mutations that may be passed forward to every new cell produced from the abnormal cell. To prevent a compromised cell from continuing to divide, there are internal control mechanisms that operate at three main **cell cycle checkpoints** at which the cell cycle can be stopped until conditions are favorable. These checkpoints occur near the end of G_1, at the G_2–M transition, and during metaphase (**Figure 12.18**).

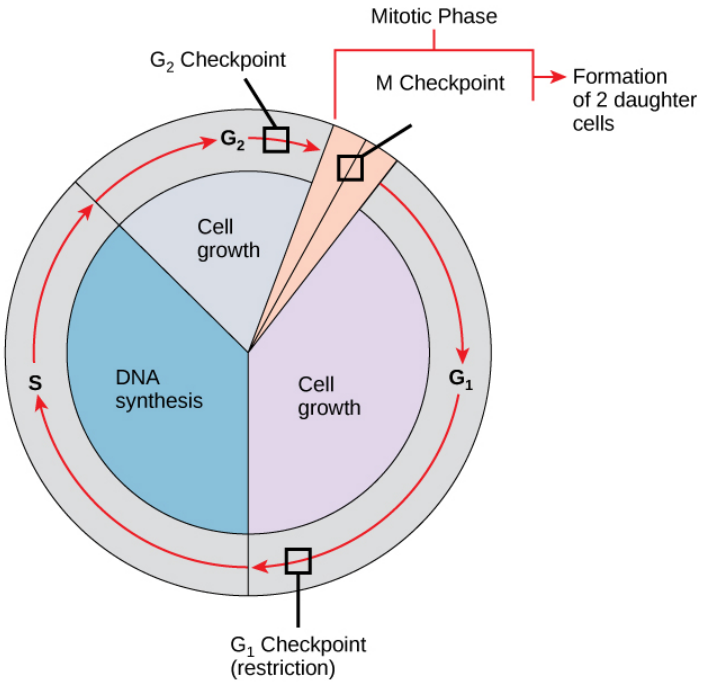


Figure 12.18 The cell cycle is controlled at three checkpoints. Integrity of the DNA is assessed at the G_1 checkpoint. Proper chromosome duplication is assessed at the G_2 checkpoint. Attachment of each kinetochore to a spindle fiber is assessed at the M checkpoint.

The G_1 Checkpoint

The G_1 checkpoint determines whether all conditions are favorable for cell division to proceed. The G_1 checkpoint, also called the restriction point, is the point at which the cell irreversibly commits to the cell-division process. In addition to adequate reserves and cell size, there is a check for damage to the genomic DNA at the G_1 checkpoint. A cell that does not meet all the requirements will not be released into the S phase.

The G_2 Checkpoint

The G_2 checkpoint bars the entry to the mitotic phase if certain conditions are not met. As in the G_1 checkpoint, cell size and protein reserves are assessed. However, the most important role of the G_2 checkpoint is to ensure that all of the chromosomes have been replicated and that the replicated DNA is not damaged.

The M Checkpoint

The M checkpoint occurs near the end of the metaphase stage of mitosis. The M checkpoint is also known as the spindle checkpoint because it determines if all the sister chromatids are correctly attached to the spindle microtubules. Because the separation of the sister chromatids during anaphase is an irreversible step, the cycle will not proceed until the kinetochores of each pair of sister chromatids are firmly anchored to spindle fibers arising from opposite poles of the cell.

12.6 | Cancer and the Cell Cycle

Introduction

> “ While there are several chronic diseases more destructive to life than cancer, none is more feared.”

Charles Mayo, 1926

Mayo's words are still true today; a diagnosis of cancer is a fearful thing. But what is cancer? Cancer is a collective name for **many different diseases** caused by a common mechanism: uncontrolled cell division. Despite the redundancy

and overlapping levels of control of cell division, errors occur. One of the critical processes monitored by the cell-cycle checkpoint surveillance mechanism is the proper replication of DNA during the S phase. Even when all of the cell-cycle controls are fully functional, a small percentage of replication errors (mutations) will be passed on to the daughter cells. If one of these changes to the DNA nucleotide sequence occurs within a gene, a gene mutation results. All cancers begin when a gene mutation gives rise to a faulty protein that participates in the process of cell reproduction. The change in the cell that results from the malformed protein may be minor. Even minor mistakes, however, may allow subsequent mistakes to occur more readily. Over and over, small, uncorrected errors are passed from parent cell to daughter cells and accumulate as each generation of cells produces more non-functional proteins from uncorrected DNA damage. Eventually, the pace of the cell cycle speeds up as the effectiveness of the control and repair mechanisms decreases. Uncontrolled growth of the mutated cells outpaces the growth of normal cells in the area, and a cancerous tumor can result.

Some definitions

All of us have heard the words cancer, tumor, malignancy, metastasis, etc. But it is important to understand the definitions of these words, and others, before we get into a discussion of the causes of the disease we know as cancer. In previous sections you learned about the cell cycle, which controls mitosis, and thus controls the growth of cells, tissues, and organs. If there is a malfunction at one of the checkpoints of the cell cycle, leading to mitosis in cells that would otherwise not divide, it would result in a population of cells which have lost control over how and when they divide. This accumulation of cells is called a **neoplasm** (from the Greek νεο- neo- "new" and πλάσμα plasma "formation, creation"). A neoplasm that forms a visible or palpable lump in the body is called a **tumor**. Tumors can be **benign**, or **malignant**, depending on how fast they grow and how readily (or not) they spread to other tissues. An example of a benign tumor would be a wart. These usually grow slowly and the cells, although they have lost cell-cycle control, do not spread to adjacent or distant tissues. A malignant neoplasm is what most people would call cancer; it grows more rapidly and can spread to adjacent or even distant sites in the body (a process known as **metastasis**. The number of blood vessels providing nutrients to the tumor may also increase (a process known as **tumor angiogenesis**).

Characteristics of cancer cells

What are the characteristics of a cancer cell, and how does it differ from a normal cell? Over the decades scientists have discovered many morphological and physiological differences (**Figure 12.19**), and studying those differences led to many of the advances in our knowledge of the cell cycle and its regulation. Cancer biologists have summarized and analyzed many of these known differences. It is known that cancer can result from mutations in many genes, and that cancers in different organs differ in their physiology, appearance, growth rate, and many other parameters. But when they filtered through all the data, they concluded that there are six essential alterations in cell physiology that are important hallmarks of the malignant state.

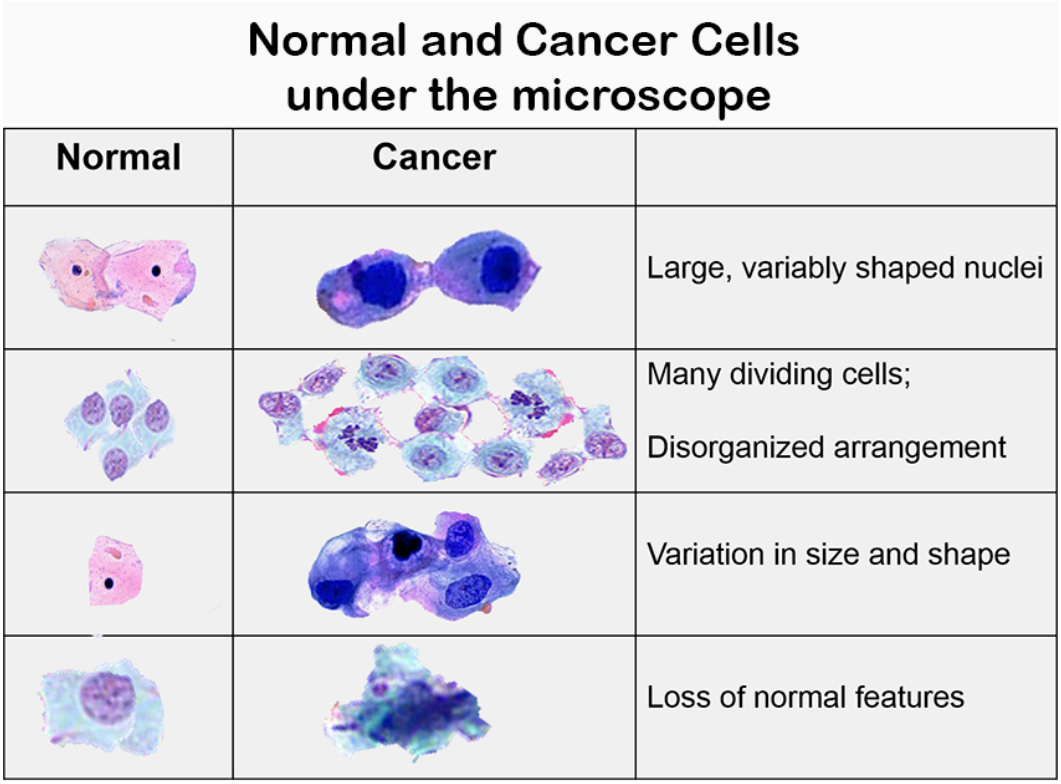


Figure 12.19 Cancer and normal cells Some characteristics of cancer cells, compared to normal cells. Figure courtesy of Dr. Wayne LaMorte, Boston University School of Public Health.

- Self-sufficiency in growth signals (positive cell-cycle regulators): Cancer cells have an unregulated ability to proliferate. This uncontrolled mitosis often occurs via the activation of **oncogenes** (literally, a gene that can cause cancer). Many of these genes code for enzymes, such as the protein kinase known as Cdk (see below) or Src, which become hyperactive when mutated; this hyperactivity drives unregulated cell proliferation (see below).
- Insensitivity to growth-inhibitory signals (negative cell cycle regulators): Cancer cells inactivate so-called tumor suppressor genes, such as RB1 or p53 (see below), that normally act at certain points in the cell cycle to inhibit mitosis (see below).
- Evasion of programmed cell death (**apoptosis**): cancer cells suppress and inactivate genes and pathways that normally cause cells to die.
- Unlimited replication potential: Cancer cells activate specific gene pathways that render them immortal even after generations of growth. HeLa cells, a human cancer cell derived from a cervical carcinoma in the 1950's, are busily proliferating in labs around the world today, long after the cancer victim passed away.
- Sustained **angiogenesis** (ability to make new blood vessels and obtain nutrients via increased blood flow): Many cancer cells acquire the capacity to induce growth of blood vessels into the tumor; this is known as tumor angiogenesis.
- Tissue invasion and metastasis: Most normal cells do not migrate, nor do they invade surrounding tissues; cancer cells acquire the capacity to migrate to other organs, invade other tissues, and colonize these organs, resulting in their spread throughout the body. This process is called **metastasis**.

Proto-oncogenes

The genes that code for the positive cell-cycle regulators are called **proto-oncogenes**. Proto-oncogenes are normal genes that, when mutated, become **oncogenes**—genes that cause a cell to become cancerous. Consider what might happen to the cell cycle in a cell with a recently acquired oncogene. In most instances, the alteration of the DNA sequence will result in a less functional (or non-functional) protein. The result is detrimental to the cell and will likely prevent the cell from completing the cell cycle; however, the organism is not harmed because the mutation will not be carried forward. If a cell cannot reproduce, the mutation is not propagated and the damage is minimal. Occasionally, however, a gene mutation causes a change that increases the activity of a positive regulator. For example, a mutation that allows Cdk, a protein involved in cell-cycle regulation, to be activated before it should be could push the cell cycle past a checkpoint before all of the required

conditions are met. If the resulting daughter cells are too damaged to undertake further cell divisions, the mutation would not be propagated and no harm comes to the organism. However, if the atypical daughter cells are able to divide further, the subsequent generation of cells will likely accumulate even more mutations, some possibly in additional genes that regulate the cell cycle.

The Cdk example is only one of many genes that are considered proto-oncogenes. In addition to the cell-cycle regulatory proteins, any protein that influences the cycle can be altered in such a way as to override cell-cycle checkpoints. Once a proto-oncogene has been altered such that there is an increase in the rate of the cell cycle, it is then called an oncogene.

Tumor Suppressor Genes

Like proto-oncogenes, many of the negative cell-cycle regulatory proteins were discovered in cells that had become cancerous. **Tumor suppressor genes** are genes that code for the negative regulator proteins, the type of regulator that—when activated—can prevent the cell from undergoing uncontrolled division. The collective function of the best-understood tumor suppressor gene proteins, retinoblastoma protein (RB1), p53, and p21, is to put up a roadblock to cell-cycle progress until certain events are completed. A cell that carries a mutated form of a negative regulator might not be able to halt the cell cycle if there is a problem.

Mutated p53 genes have been identified in more than half of all human tumor cells. This discovery is not surprising in light of the multiple roles that the p53 protein plays at the G_1 checkpoint. The p53 protein activates other genes whose products halt the cell cycle (allowing time for DNA repair), activates genes whose products participate in DNA repair, or activates genes that initiate cell death when DNA damage cannot be repaired. A damaged p53 gene can result in the cell behaving as if there are no mutations (**Figure 12.20**). This allows cells to divide, propagating the mutation in daughter cells and allowing the accumulation of new mutations. In addition, the damaged version of p53 found in cancer cells cannot trigger cell death.

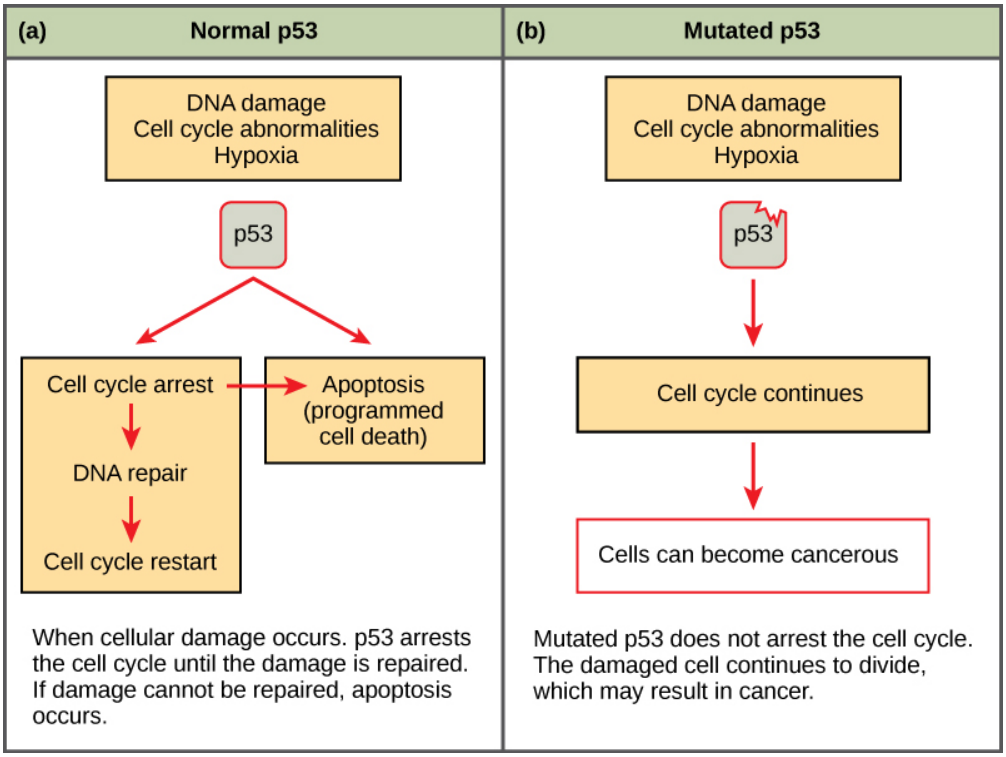


Figure 12.20 (a) The role of p53 is to monitor DNA. If damage is detected, p53 triggers repair mechanisms. If repairs are unsuccessful, p53 signals apoptosis. (b) A cell with an abnormal p53 protein cannot repair damaged DNA and cannot signal apoptosis. Cells with abnormal p53 can become cancerous. (credit: modification of work by Thierry Soussi)

13 | MOLECULAR GENETICS:DNA TO PROTEIN TO PHENOTYPE

13.1 | Transcription

Introduction

> “My own thinking (and that of many of my colleagues) is based on two general principles, which I shall call the Sequence Hypothesis and the Central Dogma. The direct evidence for both of them is negligible, but I have found them to be of great help in getting to grips with these very complex problems. I present them here in the hope that others can make similar use of them. Their speculative nature is emphasized by their names. It is an instructive exercise to attempt to build a useful theory without using them. One generally ends in the wilderness. **The Sequence Hypothesis** This has already been referred to a number of times. In its simplest form it assumes that the specificity of a piece of nucleic acid is expressed solely by the sequence of its bases, and that this sequence is a (simple) code for the amino acid sequence of a particular protein... **The Central Dogma** This states that once 'information' has passed into protein it cannot get out again. In more detail, the transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible. Information means here the precise determination of sequence, either of bases in the nucleic acid or of amino acid residues in the protein. This is by no means universally held—Sir Macfarlane Burnet, for example, does not subscribe to it—but many workers now think along these lines. As far as I know it has not been explicitly stated before. ”

Francis Crick, 1958

Crick's concept of the Central Dogma of Molecular Biology is a useful way to summarize a lot of what we know about nucleic acids. In both prokaryotes and eukaryotes, the second function of DNA (the first was replication) is to provide the

information needed to construct the proteins necessary so that the cell can perform all of its functions. To do this, the DNA is "read" or transcribed into an **mRNA** molecule. The mRNA then provides the code to form a protein by a process called translation. Through the processes of transcription and translation, a protein is built with a specific sequence of amino acids that was originally encoded in the DNA. This section discusses the details of transcription.

The Central Dogma of Molecular Biology: DNA Encodes RNA; RNA Encodes Protein

The flow of genetic information in cells from DNA to mRNA to protein is described by the **Central Dogma of Molecular Biology** (Figure 13.1), which states that genes specify the sequences of mRNAs, which in turn specify the sequences of proteins.

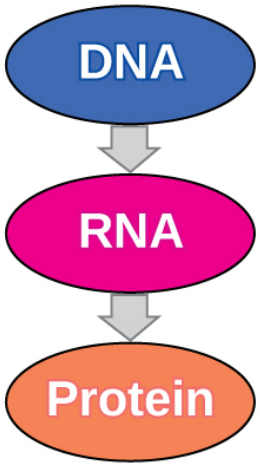


Figure 13.1 The central dogma of molecular biology states that DNA encodes RNA, which in turn encodes protein.

The copying of DNA to mRNA is relatively straightforward, with one nucleotide being added to the mRNA strand for every complementary nucleotide read in the DNA strand. The translation to protein is more complex because groups of three mRNA nucleotides correspond to one amino acid of the protein sequence. However, as we shall see in the next section, the translation to protein is still systematic, such that nucleotides 1 to 3 correspond to amino acid 1, nucleotides 4 to 6 correspond to amino acid 2, and so on.

Transcription: from DNA to mRNA

Both prokaryotes and eukaryotes perform fundamentally the same process of transcription, with the important difference of the membrane-bound nucleus in eukaryotes. With the genes bound in the nucleus, transcription occurs in the nucleus of the cell and the mRNA transcript must be transported to the cytoplasm. The prokaryotes, which include bacteria and archaea, lack membrane-bound nuclei and other organelles, and transcription occurs in the cytoplasm of the cell. In both prokaryotes and eukaryotes, transcription occurs in three main stages: initiation, elongation, and termination.

Initiation

Transcription requires the DNA double helix to partially unwind in the region of mRNA synthesis. The region of unwinding is called a transcription bubble. The DNA sequence onto which the proteins and enzymes involved in transcription bind to initiate the process is called a promoter. In most cases, promoters exist upstream of the genes they regulate. The specific sequence of a promoter is very important because it determines whether the corresponding gene is transcribed all of the time, some of the time, or hardly at all (Figure 13.2).

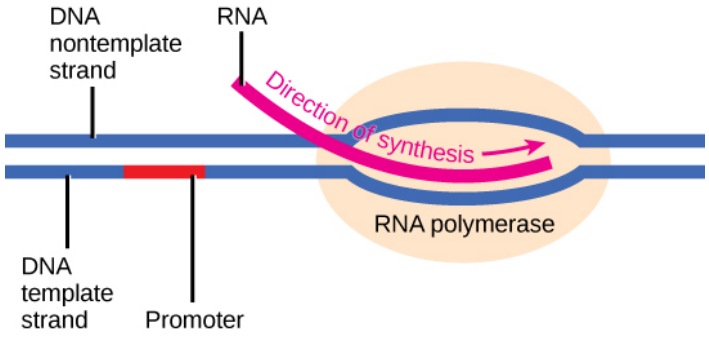


Figure 13.2 The initiation of transcription begins when DNA is unwound, forming a transcription bubble. Enzymes and other proteins involved in transcription bind at the promoter.

Elongation

Transcription always proceeds from one of the two DNA strands, which is called the **template strand**. The mRNA product is complementary to the template strand and is almost identical to the other DNA strand, called the **nontemplate strand**, with the exception that RNA contains a uracil (U) in place of the thymine (T) found in DNA. During elongation, an enzyme called **RNA polymerase** proceeds along the DNA template adding nucleotides by base pairing with the DNA template in a manner similar to DNA replication, with the difference that an RNA strand is being synthesized that does not remain bound to the DNA template. As elongation proceeds, the DNA is continuously unwound ahead of the core enzyme and rewound behind it (**Figure 13.3**).

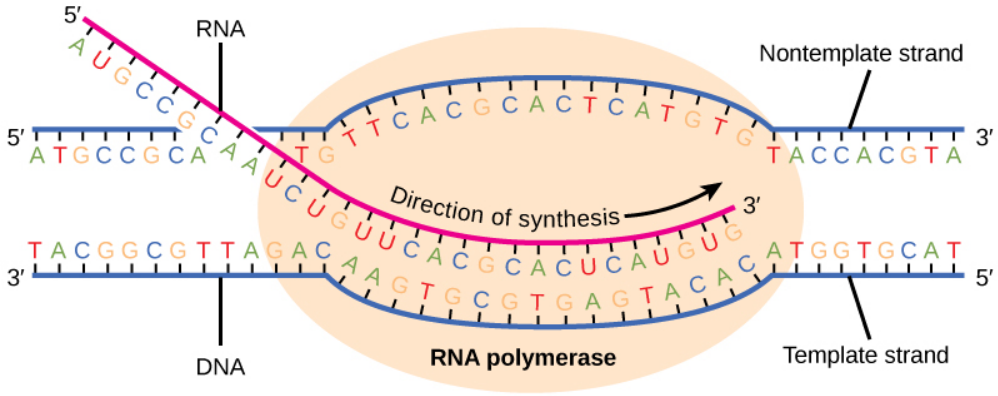


Figure 13.3 During elongation, RNA polymerase tracks along the DNA template, synthesizes mRNA in the 5' to 3' direction, and unwinds then rewinds the DNA as it is read.

Termination

Once a gene is transcribed, the prokaryotic polymerase needs to be instructed to dissociate from the DNA template and liberate the newly made mRNA. Depending on the gene being transcribed, there are two kinds of termination signals, but both involve repeated nucleotide sequences in the DNA template that result in RNA polymerase stalling, leaving the DNA template, and freeing the mRNA transcript.

On termination, the process of transcription is complete. In a prokaryotic cell, by the time termination occurs, the transcript would already have been used to partially synthesize numerous copies of the encoded protein because these processes can occur concurrently using multiple ribosomes (polyribosomes) (**Figure 13.4**). In contrast, the presence of a nucleus in eukaryotic cells precludes simultaneous transcription and translation.

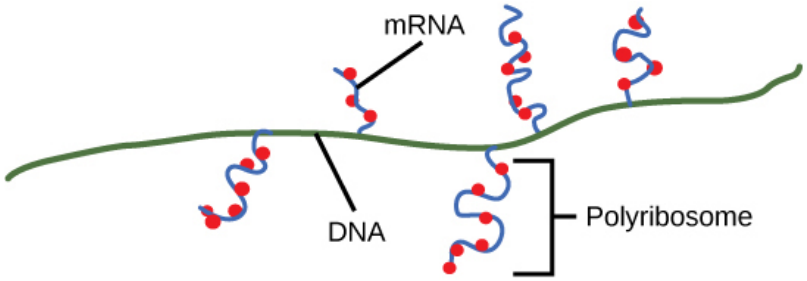


Figure 13.4 Multiple polymerases can transcribe a single bacterial gene while numerous ribosomes concurrently translate the mRNA transcripts into polypeptides. In this way, a specific protein can rapidly reach a high concentration in the bacterial cell.

Eukaryotic RNA Processing

The newly transcribed eukaryotic mRNAs must undergo several processing steps before they can be transferred from the nucleus to the cytoplasm and translated into a protein. The additional steps involved in eukaryotic mRNA maturation create a molecule that is much more stable than a prokaryotic mRNA. For example, eukaryotic mRNAs last for several hours, whereas the typical prokaryotic mRNA lasts no more than five seconds.

The mRNA transcript is first coated in RNA-stabilizing proteins to prevent it from degrading while it is processed and exported out of the nucleus. This occurs while the pre-mRNA still is being synthesized by adding a special nucleotide "cap" to the 5' end of the growing transcript. In addition to preventing degradation, factors involved in protein synthesis recognize the cap to help initiate translation by ribosomes.

Once elongation is complete, an enzyme then adds a string of approximately 200 adenine residues to the 3' end, called the poly-A tail. This modification further protects the pre-mRNA from degradation and signals to cellular factors that the transcript needs to be exported to the cytoplasm.

Eukaryotic genes are composed of protein-coding sequences called exons (*ex*-on signifies that they are *ex*pressed) and *int*ervening sequences called introns (*int*-ron denotes their *int*ervening role). Introns are removed from the pre-mRNA during processing. Intron sequences in mRNA do not encode functional proteins. It is essential that all of a pre-mRNA's introns be completely and precisely removed before protein synthesis so that the exons join together to code for the correct amino acids. If the process errs by even a single nucleotide, the sequence of the rejoined exons would be shifted, and the resulting protein would be nonfunctional. The process of removing introns and reconnecting exons is called splicing (**Figure 13.5**). Introns are removed and degraded while the pre-mRNA is still in the nucleus.

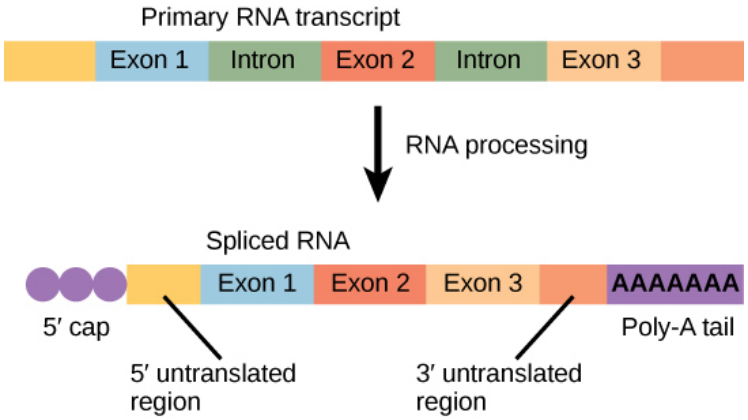


Figure 13.5 Eukaryotic mRNA contains introns that must be spliced out. A 5' cap and 3' tail are also added.

13.2 | Translation

Introduction

> “ Protein synthesis is a central problem for the whole of biology, and that it is in all probability closely related to gene action. ”

Francis Crick, 1958

Crick's words are still true today. The synthesis of proteins is one of a cell’s most complicated and energy-consuming metabolic processes, and the evolution of this complex process remains one of the hardest questions to answer. But the products of protein synthesis (proteins) indeed are closely related to gene action. In addition, proteins account for more mass than any other component of living organisms (with the exception of water), and proteins perform a wide variety of the functions of a cell. The process of translation, or protein synthesis, involves decoding an mRNA message into a polypeptide product, in which amino acids are covalently strung together in lengths ranging from approximately 50 amino acids to more than 1,000.

The Protein Synthesis Machinery

In addition to the mRNA template, many other molecules contribute to the process of translation. The composition of each component may vary across species; for instance, ribosomes may consist of different numbers of ribosomal RNAs (**rRNA**) and polypeptides depending on the organism. However, the general structures and functions of the protein synthesis machinery are comparable from bacteria to human cells. Translation requires the input of an mRNA template, ribosomes, tRNAs, and various enzymatic factors (**Figure 13.6**).

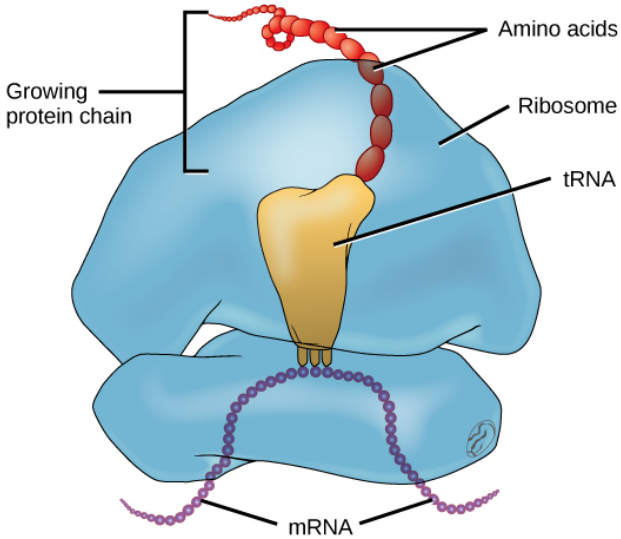


Figure 13.6 The protein synthesis machinery includes the large and small subunits of the ribosome, mRNA, and tRNA. (credit: modification of work by NIGMS, NIH)

In *E. coli*, there are 200,000 ribosomes present in every cell at any given time. A ribosome is a complex macromolecule composed of structural and catalytic rRNAs, and many distinct polypeptides. In eukaryotes, the nucleolus is completely specialized for the synthesis and assembly of rRNAs.

Ribosomes are located in the cytoplasm and on the plasma membrane of prokaryotes, and in the cytoplasm and endoplasmic reticulum of eukaryotes. Ribosomes are made up of a large and a small subunit that come together for translation. The small subunit is responsible for binding the mRNA template, whereas the large subunit sequentially binds **tRNAs**, a type of RNA molecule that brings amino acids to the growing chain of the polypeptide. Each mRNA molecule is simultaneously translated by many ribosomes, all synthesizing protein in the same direction.

Depending on the species, 40 to 60 types of tRNA exist in the cytoplasm. Serving as adaptors, specific tRNAs bind to sequences on the mRNA template and add the corresponding amino acid to the polypeptide chain. Therefore, tRNAs are the molecules that actually “translate” the language of RNA into the language of proteins. For each tRNA to function, it must have its specific amino acid bonded to it. In the process of tRNA “charging,” each tRNA molecule is bonded to its correct amino acid.

The Genetic Code

To summarize what we know to this point, the cellular process of transcription generates messenger RNA (mRNA), a mobile molecular copy of one or more genes with an alphabet of A, C, G, and uracil (U). Translation of the mRNA template converts nucleotide-based genetic information into a protein product. Protein sequences consist of 20 commonly occurring amino acids; therefore, it can be said that the protein alphabet consists of 20 letters. Each amino acid is defined by a three-nucleotide sequence called the triplet **codon**. The relationship between a nucleotide codon and its corresponding amino acid is called the **genetic code**.

Given the different numbers of "letters" in the mRNA and protein "alphabets," combinations of nucleotides corresponded to single amino acids. Using a three-nucleotide code means that there are a total of 64 (4 × 4 × 4) possible combinations; therefore, a given amino acid is encoded by more than one nucleotide triplet (**Figure 13.7**).

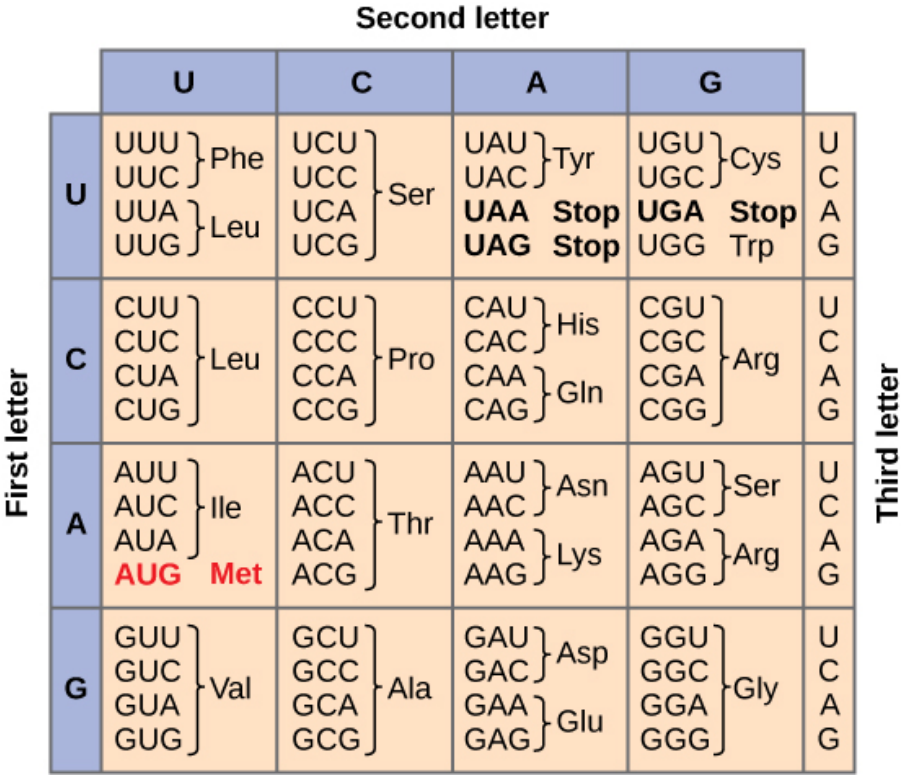


Figure 13.7 This figure shows the genetic code for translating each nucleotide triplet, or codon, in mRNA into an amino acid or a termination signal in a nascent protein. (credit: modification of work by NIH)

Three of the 64 codons terminate protein synthesis and release the polypeptide from the translation machinery. These triplets are called **stop codons**. Another codon, AUG, also has a special function. In addition to specifying the amino acid methionine, it also serves as the **start codon** to initiate translation. The reading frame for translation is set by the AUG start codon near the 5' end of the mRNA. The genetic code is universal. With a few exceptions, virtually all species use the same genetic code for protein synthesis, which is powerful evidence that all life on Earth shares a common origin.

The Mechanism of Protein Synthesis

Just as with mRNA synthesis, protein synthesis can be divided into three phases: initiation, elongation, and termination. The process of translation is similar in prokaryotes and eukaryotes. Here we will explore how translation occurs in *E. coli*, a representative prokaryote, and specify any differences between prokaryotic and eukaryotic translation.

Protein synthesis begins with the formation of an initiation complex. In *E. coli*, this complex involves the small ribosome subunit, the mRNA template, three initiation factors, and a special initiator tRNA. The initiator tRNA interacts with the AUG start codon, and links to a special form of the amino acid methionine that is typically removed from the polypeptide after translation is complete.

In prokaryotes and eukaryotes, the basics of polypeptide elongation are the same, so we will review elongation from the perspective of *E. coli*. The large ribosomal subunit of *E. coli* consists of three compartments: the A site binds incoming charged tRNAs (tRNAs with their attached specific amino acids). The P site binds charged tRNAs carrying amino acids that

have formed bonds with the growing polypeptide chain but have not yet dissociated from their corresponding tRNA. The E site releases dissociated tRNAs so they can be recharged with free amino acids. The ribosome shifts one codon at a time, catalyzing each process that occurs in the three sites. With each step, a charged tRNA enters the complex, the polypeptide becomes one amino acid longer, and an uncharged tRNA departs. The energy for each bond between amino acids is derived from GTP, a molecule similar to ATP (**Figure 13.8**). Amazingly, the *E. coli* translation apparatus takes only 0.05 seconds to add each amino acid, meaning that a 200-amino acid polypeptide could be translated in just 10 seconds.

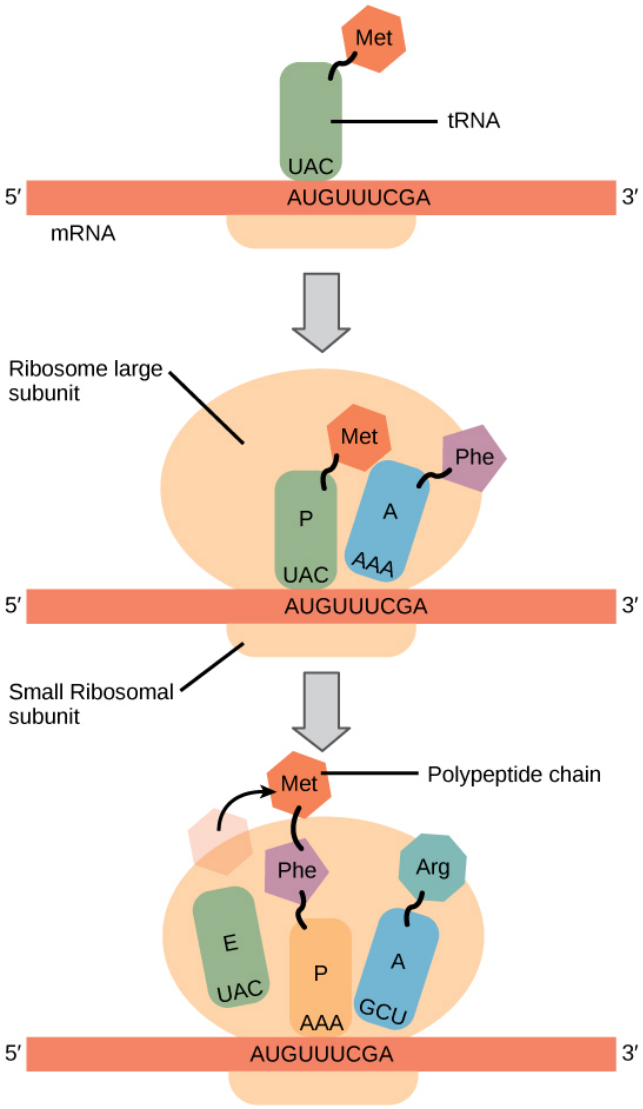


Figure 13.8 Translation begins when a tRNA anticodon recognizes a codon on the mRNA. The large ribosomal subunit joins the small subunit, and a second tRNA is recruited. As the mRNA moves relative to the ribosome, the polypeptide chain is formed. Entry of a release factor into the A site terminates translation and the components dissociate.

Termination of translation occurs when a stop codon (UAA, UAG, or UGA) is encountered. When the ribosome encounters the stop codon, the growing polypeptide is released and the ribosome subunits dissociate and leave the mRNA. After many ribosomes have completed translation, the mRNA is degraded so the nucleotides can be reused in another transcription reaction.

13.3 | Connection Between DNA and Phenotype

You CAN get there from here...

- In the beginning was the word

- WORD
- WORE
- GORE
- GONE
- GENE
- and by the mutation came the gene.

Michael A. Arbib, *Towards a Theoretical Biology: An IUBS Symposium*, 1969

Sickle Cell Anemia: A look at the connection between DNA and Phenotype

Changes in the DNA sequence of a gene are known as **mutations**. Because genes code for and are translated into proteins, mutations often (but not always) result in changes in the sequence of amino acids in those proteins. Changes in the amino acid sequence can modify or even completely destroy protein function. Proteins have many functions within cells, and a change in those functions results in a change in the phenotype of that cell or organism. So a mutation as simple as a single base change in a DNA sequence can have dramatic effects on phenotype. One of the best examples of this phenomenon can be observed when mutations occur in the gene for one of the protein components of the red blood cell protein we call hemoglobin.

A major component of the **erythrocytes** (red blood cells) found in vertebrates is **hemoglobin**. A molecule of hemoglobin from a normal adult human contains 4 proteins (two identical alpha polypeptides and two identical beta polypeptides) surrounding a core of heme (contains complex molecule containing an atom of iron which can combine reversibly with oxygen). Thus, hemoglobin functions as the major oxygen-carrying constituent of blood. Because of hemoglobin, a given volume of blood can carry far more oxygen than could be dissolved in an equal volume of water.

The polypeptides that make up the hemoglobin are produced by the processes of transcription and translation. So, the instructions for making the polypeptides are stored in the DNA, and the specific region on the DNA that codes for the polypeptide is called a **gene**. The physical location of the gene on the chromosome is the **locus**. If you remember from the earlier section on protein structure and function, the sequence of amino acids in a polypeptide can have a significant effect on the overall function of the protein. Since the sequence of nucleotides in the DNA codes for a particular sequence of amino acids in the polypeptide, it is logical to conclude that changes in the nucleotide sequence in the DNA can change the sequence of amino acids in the polypeptide and thereby changing how the polypeptide functions.

In many human populations, particularly those with origins in Central Africa or the Mediterranean, there are individuals who suffer from severe anemia and whose blood contains numerous distorted, sickle-shaped erythrocytes (**Figure 13.9**). Hence, the disease was given the name sickle-cell anemia.

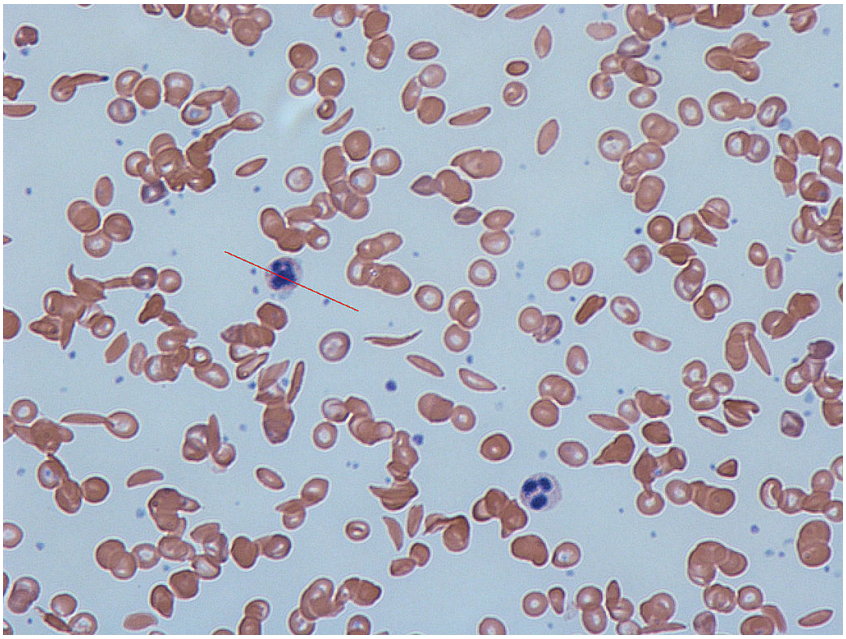


Figure 13.9 Notice the sickle shaped cells in the image. By Dr Graham Beards via Wikimedia Commons

Sickle-cell anemia is one of the most studied and well-understood genetic diseases. In spite of this level of understanding, there is currently no effective treatment for the disease. Experimental bone marrow transplants have cured the disease in some patients, but bone marrow transplants are currently recommended in only a handful of cases. In the United States alone, there are 50,000 individuals with sickle-cell anemia.

Biochemical studies established that the gene affected in sickle-cell anemia has the code for an abnormal beta polypeptide, which is one of the components of the hemoglobin molecule. Therefore, there are two different forms of the hemoglobin gene that codes for the beta chain:

- Form 1: $\mathbf{Hb^A}$ has the code for a normal beta chain
- Form 2: $\mathbf{Hb^S}$ has the code for an abnormal beta chain

$\mathbf{Hb^A}$ and $\mathbf{Hb^S}$ are considered alleles. The word "allele" comes from the same root as "alternative." Thus, alleles are alternative forms of the same gene. Alleles arise by mutation. Mutations come about by many different ways, but what is common among the different types of mutations is a change in the order of nucleotides in the DNA. Thinking back to the Central Dogma of Molecular Biology, one can reword it with the above in mind. Changes in the DNA causes changes in RNA which can cause changes in the polypeptide and a new polypeptide is a new phenotype.

Humans are **diploid** organisms; they have two copies of most genes. However, the two copies they possess do not have to be identical. When there are two possible alleles for a gene (such as in the gene for the beta chain of hemoglobin), a diploid individual will have one of three possible combinations of the two alleles. They can be $\mathbf{Hb^A\ Hb^A}$, $\mathbf{Hb^A\ Hb^S}$, or $\mathbf{Hb^S\ Hb^S}$.

The set of alleles present in an individual for a given gene is known as the individual's **genotype**. The three combinations of two alleles above are therefore the three different genotypes. Individuals that have two copies of the same allele are called **homozygous**; individuals with two different alleles are called **heterozygous**. So, an individual that is $\mathbf{Hb^A\ Hb^A}$ is homozygous normal beta chain, an individual that is $\mathbf{Hb^A\ Hb^S}$ is heterozygous, and an individual that is $\mathbf{Hb^S\ Hb^S}$ is homozygous abnormal beta chain.

There is one more term that you need to know: the expression of the particular combination of alleles an individual has at a locus is known as the individual's phenotype. The **phenotype** is really nothing more than the final end product of the genes and their interaction with the environment.

Homozygous $\mathbf{Hb^S\ Hb^S}$ individuals (called sickle-cell anemics) can have many sickle-shaped (as opposed to normal disc-shaped) erythrocytes in their blood. How does this occur? The process is well-understood, and summarized below.

In the capillaries (microscopic blood vessels that directly exchange oxygen with the tissues), erythrocytes can be subjected to low oxygen tension after they lose their oxygen to the surrounding tissues. In this low oxygen situation, the abnormal hemoglobin molecules of $\mathbf{Hb^S\ Hb^S}$ individuals tend to polymerize (join together), forming stiff, tubular fibers which

ultimately distort the shape of the entire erythrocyte, giving it the characteristic “sickle” shape. These sickled cells have a number of effects on the body via two processes.

- 1. The sickle-shape affects capillary blood flow:
- Sickled cells are less able to enter and move through the capillaries.
- Once in the capillaries, they clog capillary flow and cause small blood clots.
- Reduced blood flow results in reduced oxygen availability to the tissues.
- Reduced oxygen supply results in tissue death and damage to vital organs (e.g., the heart, liver and spleen).
- 2. Sickled blood cells have a shorter lifespan than normal red blood cells:
- Reduced lifespan of erythrocytes places a greater demand on the bone marrow to make new red blood cells and on the spleen to break down dead erythrocytes.
- Increased demand on the bone marrow results in severe pain in the long bones and joints.
- **Individuals suffering from sickle-cell anemia are frequently ill and generally have a considerably reduced lifespan. These individuals are said to have sickle-cell disease.**

Homozygous $\mathbf{Hb^A\ Hb^A}$ individuals have erythrocytes which retain their normal shape in the body and which retain normal shape even when blood samples are subjected to greatly reduced oxygen tension in laboratory tests.

Heterozygous individuals ($\mathbf{Hb^A\ Hb^S}$) are said to be **carriers** for sickle-cell anemia. Note: this is a specific term and is not the same thing as sickle cell anemia—heterozygotes do not have the disease themselves but their children may inherit the condition. Carriers have no anemia, do have good health (as do $\mathbf{Hb^A\ Hb^A}$ individuals) and their erythrocytes maintain normal shape in the blood. In other words, they are phenotypically normal under most conditions, and probably do not know that they “carry” the $\mathbf{Hb^S}$ allele. However, if heterozygotes are exposed to conditions of low oxygen levels (such as strenuous activity at high altitudes) some of their erythrocytes do sickle. Red blood cells in blood samples of heterozygotes subjected to greatly reduced oxygen tension in the laboratory also sickle.

Why is sickle cell anemia most prevalent in people with origins in Central Africa and the Mediterranean? If you look at the figure below (**Figure 13.10**) , you will see the occurrence of sickle cell anemia overlaps with the pervasiveness of malaria. This seems odd, but those individuals how are heterozygous ($\mathbf{Hb^A\ Hb^S}$) for the sickle cell allele are less likely to contract and die from malaria then those who are homozygous ($\mathbf{Hb^A\ Hb^A}$). The $\mathbf{Hb^S}$ polypeptide that is produced by the heterozygous individual stops the organism (*Plasmodium*) that causes malaria from invading the red blood cells. So, in areas where malaria is common there is selection pressure for the $\mathbf{Hb^S}$ allele, and the $\mathbf{Hb^S}$ allele occurs in a higher frequency because the those who have one copy of the $\mathbf{Hb^S}$ allele will live longer and have more children. In areas where malaria is not common, there is selection pressure against the $\mathbf{Hb^S}$ allele, and the $\mathbf{Hb^S}$ allele occurs in a lower frequency. As you will learn in a later chapter, there is an 25% chance that two carriers will have a child who is homozygous $\mathbf{Hb^S\ Hb^S}$), and this child will pay the evolutionary price for the protection from malaria that the parents were afforded. Hopefully, you will now understand how evolution favors the presence of such a potentially detrimental allele in a population. The sickle cell example is only one of what is called heterozygous advantage, we have provided a number of other examples in **Table 13.1**.

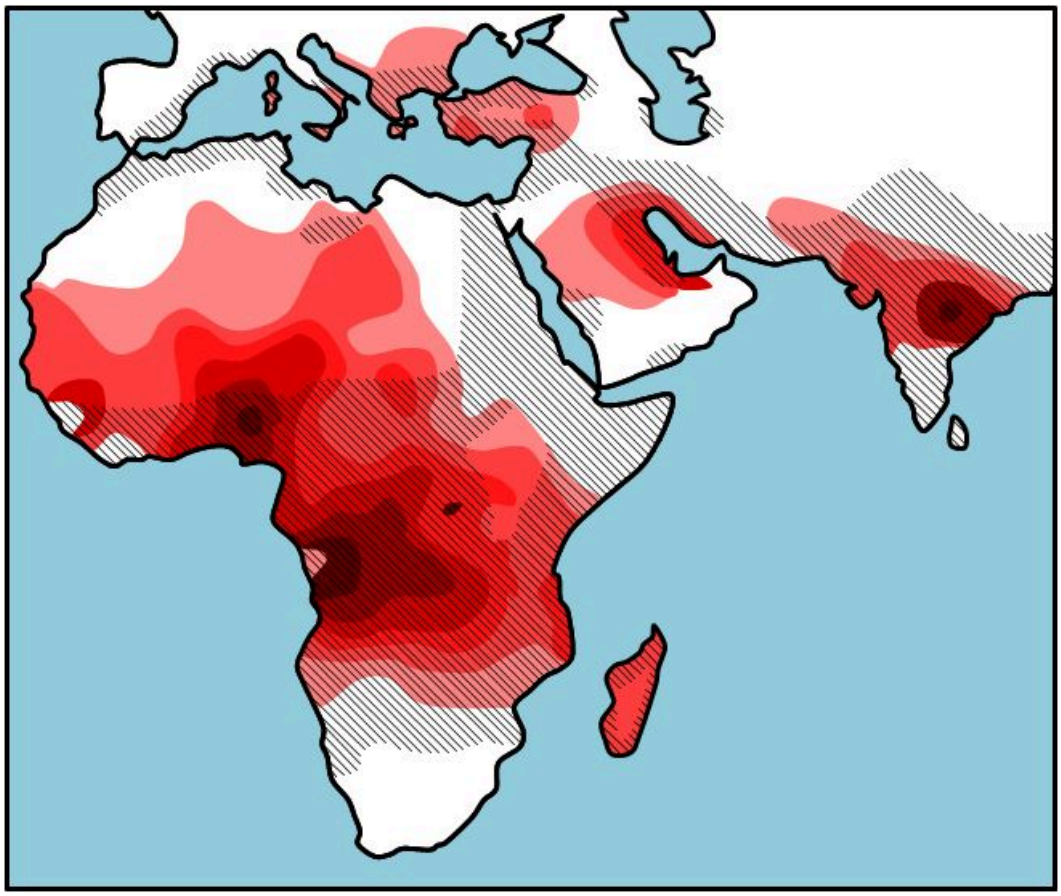


Figure 13.10 Distribution of malaria and the frequency of sickle cell allele The hatched line represents the distribution of malaria. The various red colors represent the relative frequency of sickle cell allele in the population with the dark red having the highest frequency and the light red having the lowest frequency. Work by Eva Horne.

Recessive Illness	Heterozygote Advantage	Possible Explanation
Cystic fibrosis	protection against diarrheal diseases such as cholera	Carriers have too few functional chloride channels in intestinal cells, blocking toxin
G6PD Deficiency	Protection against malaria	Red blood cells inhospitable to malaria
Phenylketonuria (PKU)	Protection against miscarriage induced by a fungal toxin	Excess amino acid (phenylalanine) in carriers inactivates toxin
Tay-Sachs disease	Protection against tuberculosis	Unknown
Noninsulin-dependent diabetes mellitus	Protection against starvation	Tendency to gain weight protects against starvation during famine

Table 13.1 Examples of Heterozygous Advantage in Humans

14 | MEIOSIS AND MENDELIAN GENETICS

14.1 | Sexual Reproduction

Introduction

> “Being the inventor of sex would seem to be a sufficient distinction for a creature just barely large enough to be seen by the naked eye. [Comment about Volvox, a freshwater green algae, which appears indetermimately plantlike and animal-like during its reproductive cycle.] ”

Joseph Wood Krutch, 1957

Sexual reproduction was an early evolutionary innovation after the appearance of eukaryotic cells. The fact that most eukaryotes reproduce sexually is evidence of its evolutionary success. In many animals, it is the only mode of reproduction. And yet, scientists recognize some real disadvantages to sexual reproduction. On the surface, offspring that are genetically identical to the parent may appear to be more advantageous. If the parent organism is successfully occupying a habitat, offspring with the same traits would be similarly successful. There is also the obvious benefit to an organism that can produce offspring by asexual budding, fragmentation, or asexual eggs. These methods of reproduction do not require another organism of the opposite sex. There is no need to expend energy finding or attracting a mate. That energy can be spent on producing more offspring. Indeed, some organisms that lead a solitary lifestyle have retained the ability to reproduce asexually. In addition, asexual populations only have female individuals, so every individual is capable of reproduction. In contrast, the males in sexual populations (half the population) are not producing offspring themselves. Because of this, an asexual population can grow twice as fast as a sexual population in theory. This means that in competition, the asexual population would have the advantage. All of these advantages to asexual reproduction, which are also disadvantages to sexual reproduction, should mean that the number of species with asexual reproduction should be more common.

However, multicellular organisms that exclusively depend on asexual reproduction are exceedingly rare. Why is sexual reproduction so common? This is one of the important questions in biology and has been the focus of much research from the latter half of the twentieth century until now. A likely explanation is that the variation that sexual reproduction creates among offspring is very important to the survival and reproduction of those offspring. The only source of variation in asexual organisms is mutation. This is the ultimate source of variation in sexual organisms. In addition, those different mutations are continually reshuffled from one generation to the next when different parents combine their unique genomes, and the genes are mixed into different combinations by the process of **meiosis**. Meiosis is the division of the contents of the nucleus that divides the chromosomes among gametes. Variation is introduced during meiosis, as well as when the gametes combine in fertilization.

evolution IN ACTION

The Red Queen Hypothesis

There is no question that sexual reproduction provides evolutionary advantages to organisms that employ this mechanism to produce offspring. The problematic question is why, even in the face of fairly stable conditions, sexual reproduction persists when it is more difficult and produces fewer offspring for individual organisms? Variation is the outcome of sexual reproduction, but why are ongoing variations necessary? Enter the Red Queen hypothesis, first proposed by Leigh Van Valen in 1973.[1] The concept was named in reference to the Red Queen's race in Lewis Carroll's book, *Through the Looking-Glass,* in which the Red Queen says one must run at full speed just to stay where one is.

All species coevolve with other organisms. For example, predators coevolve with their prey, and parasites coevolve with their hosts. A remarkable example of coevolution between predators and their prey is the unique coadaptation of night flying bats and their moth prey. Bats find their prey by emitting high-pitched clicks, but moths have evolved simple ears to hear these clicks so they can avoid the bats. The moths have also adapted behaviors, such as flying away from the bat when they first hear it, or dropping suddenly to the ground when the bat is upon them. Bats have evolved "quiet" clicks in an attempt to evade the moth's hearing. Some moths have evolved the ability to respond to the bats' clicks with their own clicks as a strategy to confuse the bats echolocation abilities.

Each tiny advantage gained by favorable variation gives a species an edge over close competitors, predators, parasites, or even prey. The only method that will allow a coevolving species to keep its own share of the resources is also to continually improve its ability to survive and produce offspring. As one species gains an advantage, other species must also develop an advantage or they will be outcompeted. No single species progresses too far ahead because genetic variation among progeny of sexual reproduction provides all species with a mechanism to produce adapted individuals. Species whose individuals cannot keep up become extinct. The Red Queen's catchphrase was, "It takes all the running you can do to stay in the same place." This is an apt description of coevolution between competing species.

Life Cycles of Sexually Reproducing Organisms

Fertilization and meiosis alternate in sexual **life cycles**. What happens between these two events depends on the organism. The process of meiosis reduces the resulting gamete's chromosome number by half. Fertilization, the joining of two haploid gametes, restores the diploid condition. The number of sets of chromosomes in a cell is called its **ploidy** level. Haploid (1n) cells contain one set of chromosomes. Cells containing two sets of chromosomes are called diploid (2n). There are three main categories of life cycles in multicellular organisms: diploid-dominant, in which the multicellular diploid stage is the most obvious life stage (and there is no multicellular haploid stage), as with most animals including humans; haploid-dominant, in which the multicellular haploid stage is the most obvious life stage (and there is no multicellular diploid stage), as with all fungi and some algae; and alternation of generations, in which the two stages, haploid and diploid, are apparent to one degree or another depending on the group, as with plants and some algae.

Nearly all animals employ a diploid-dominant life-cycle strategy in which the only haploid cells produced by the organism are the gametes. The gametes are produced from diploid **germ cells**, a special cell line that only produces gametes. Once the haploid gametes are formed, they lose the ability to divide again. There is no multicellular haploid life stage. Fertilization occurs with the fusion of two gametes, usually from different individuals, restoring the diploid state (**Figure 14.1a**).

1. Leigh Van Valen, "A new evolutionary law," Evolutionary Theory 1 (1973): 1–30.

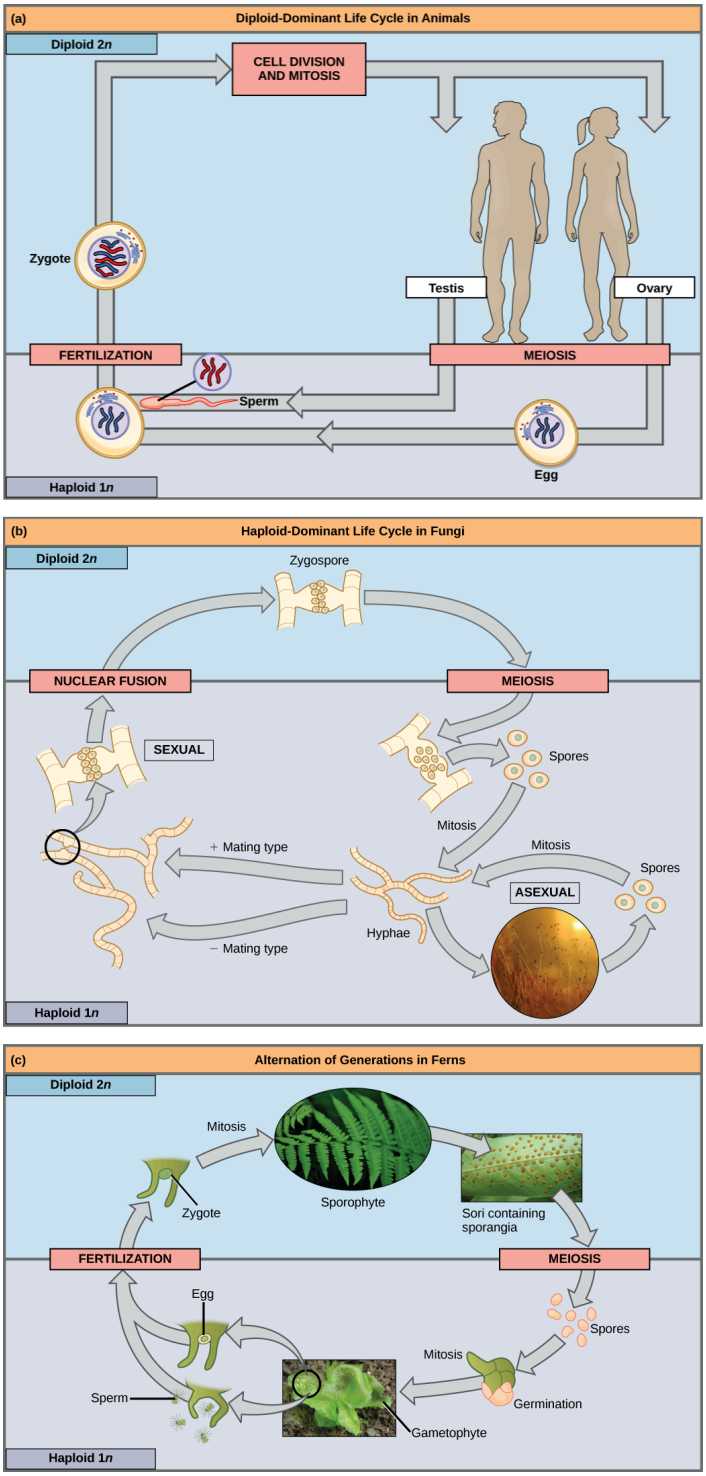


Figure 14.1 (a) In animals, sexually reproducing adults form haploid gametes from diploid germ cells. (b) Fungi, such as black bread mold (*Rhizopus nigricans*), have haploid-dominant life cycles. (c) Plants have a life cycle that alternates between a multicellular haploid organism and a multicellular diploid organism. (credit c "fern": modification of work by Cory Zanker; credit c "gametophyte": modification of work by "Vlmastra"/Wikimedia Commons)

Most fungi and algae employ a life-cycle strategy in which the multicellular "body" of the organism is haploid. During sexual reproduction, specialized haploid cells from two individuals join to form a diploid zygote. The zygote immediately undergoes meiosis to form four haploid cells called spores (**Figure 14.1b**).

The third life-cycle type, employed by some algae and all plants, is called alternation of generations. These species have both haploid and diploid multicellular organisms as part of their life cycle. The haploid multicellular plants are called gametophytes because they produce gametes. Meiosis is not involved in the production of gametes in this case, as the organism that produces gametes is already haploid. Fertilization between the gametes forms a diploid zygote. The zygote will undergo many rounds of mitosis and give rise to a diploid multicellular plant called a sporophyte. Specialized cells of the sporophyte will undergo meiosis and produce haploid spores. The spores will develop into the gametophytes (**Figure 14.1c**).

14.2 | Meiosis

Introduction

> "I may finally call attention to the probability that the association of the paternal and maternal chromosomes in pairs, and their subsequent separation during the reducing division as indicated above, may constitute the physical basis of the Mendelian law of heredity."

Walter S. Sutton, "On the Morphology of the Chromosome Group in Brachystola magna",*Biological Bulletin*, 1902, 4:39

Sutton was one of the biologists (along with Theodor Boveri) who discovered that chromosomes correspond to the paired "particles" required by Mendel's Laws, and that the number of these chromosomes is halved at the time that sperm and egg cells are generated. The Boveri-Sutton Chromosome Theory (also known as the Chromosome Theory of Inheritance) is one of the important linkages between Mendel and modern molecular biology. Sutton was raised on a farm near Russell, Kansas, and played basketball under Coach Naismith at the University of Kansas. His ground-breaking work was done at Columbia University in New York City. So when you study meiosis and heredity, remember that a major part of this knowledge came from the work of another biologist from Kansas, Walter Sutton, working with grasshoppers (the aforementioned Brachystola magna) that are indigenous to the state of Kansas.

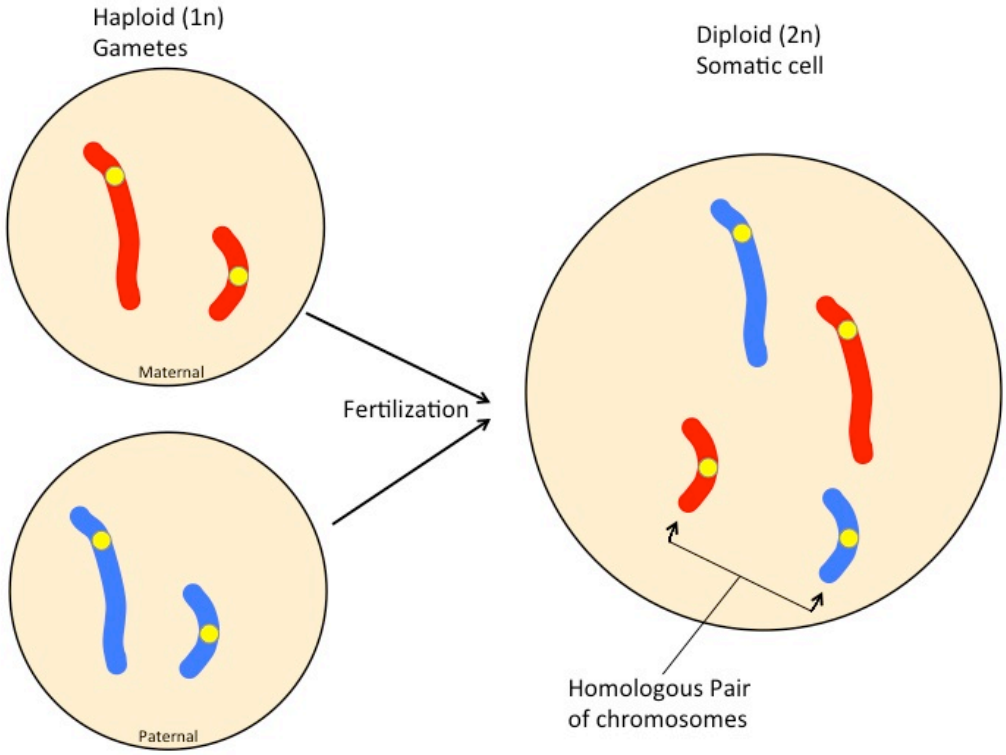


Figure 14.2 The process of fertilization produces a dipliod cell by the union of two haploid cells. The homologous pairs are based on size and the color refers to the maternal and paternal chromosomes. Work by Robbie Bear.

Sexual reproduction requires **fertilization**, a union of two cells from two individual organisms. If those two cells each contain one set of chromosomes, then the resulting cell contains two sets of chromosomes. The number of sets of chromosomes in a cell is called its **ploidy** level. Haploid (1n) cells contain one set of chromosomes. Cells containing two sets of chromosomes are called diploid (2n) (**Figure 14.2**). If the reproductive cycle is to continue, the diploid cell must somehow reduce its number of chromosome sets before fertilization can occur again, or there will be a continual doubling in the number of chromosome sets in every generation. So, in addition to fertilization, sexual reproduction includes a nuclear division, known as meiosis, that reduces the number of chromosome sets.

Most animals and plants are diploid, containing two sets of chromosomes; in each **somatic cell** (the nonreproductive cells of a multicellular organism), the nucleus contains two copies of each chromosome that are referred to as **homologous chromosomes** (**Figure 14.2**). Somatic cells are sometimes referred to as “body” cells. Homologous chromosomes are matched pairs containing genes for the same traits in identical locations along their length. Diploid organisms inherit one copy of each homologous chromosome from each parent; all together, they are considered a full set of chromosomes. In animals, haploid cells containing a single copy of each homologous chromosome are found only within gametes. Gametes fuse with another haploid gamete to produce a diploid cell.

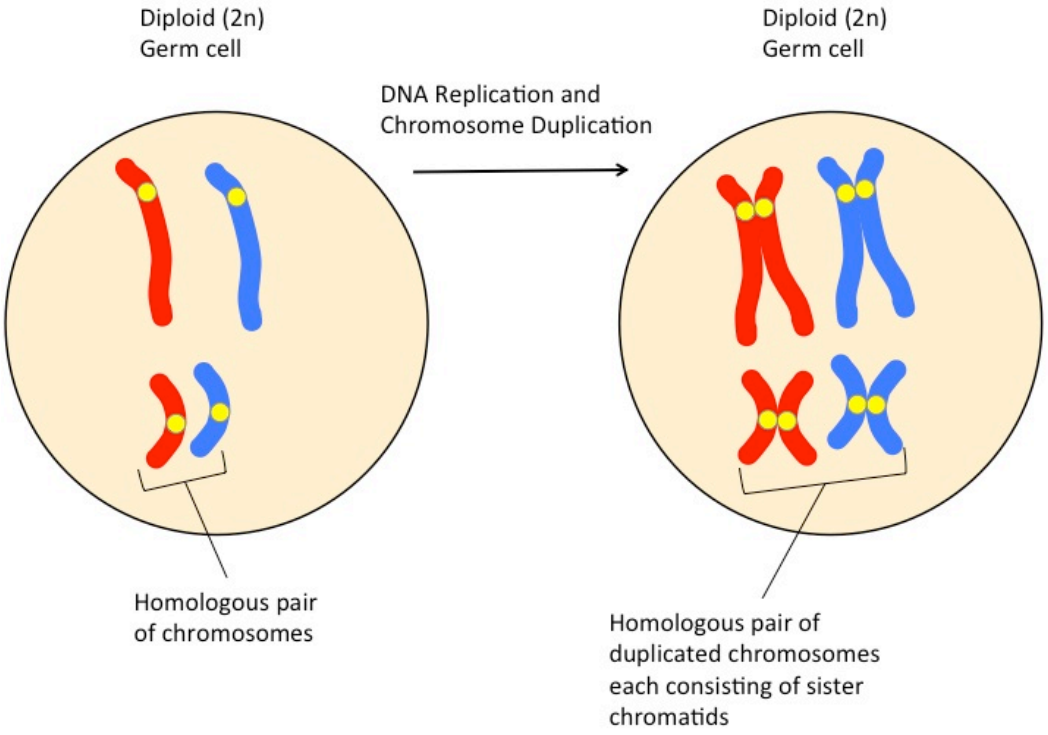


Figure 14.3 The process of DNA replication produces a duplicated chromosome with two sister chromatids in each of the homologous pairs. The homologous pairs are based on size and the color refers to the maternal and paternal chromosomes. Work by Robbie Bear.

The nuclear division that forms haploid cells, which is called meiosis, is related to mitosis. As you have learned, mitosis is part of a cell reproduction cycle that results in identical daughter nuclei that are also genetically identical to the original parent nucleus. In mitosis, both the parent and the daughter nuclei contain the same number of chromosome sets—diploid for most plants and animals. Meiosis employs many of the same mechanisms as mitosis. However, the starting nucleus is always diploid and the nuclei that result at the end of a meiotic cell division are haploid. To achieve the reduction in chromosome number, meiosis consists of one round of chromosome duplication (**Figure 14.3**) and two rounds of nuclear division. Because the events that occur during each of the division stages are analogous to the events of mitosis, the same stage names are assigned. However, because there are two rounds of division, the stages are designated with a "I" or "II." Thus, **meiosis I** is the first round of meiotic division and consists of prophase I, metaphase I, and so on. Meiosis I reduces the number of chromosome sets from two to one. The genetic information is also mixed during this division to create unique recombinant chromosomes. **Meiosis II**, in which the second round of meiotic division takes place in a way that is similar to mitosis, includes prophase II, pmetaphase II, and so on (**Figure 14.4**).

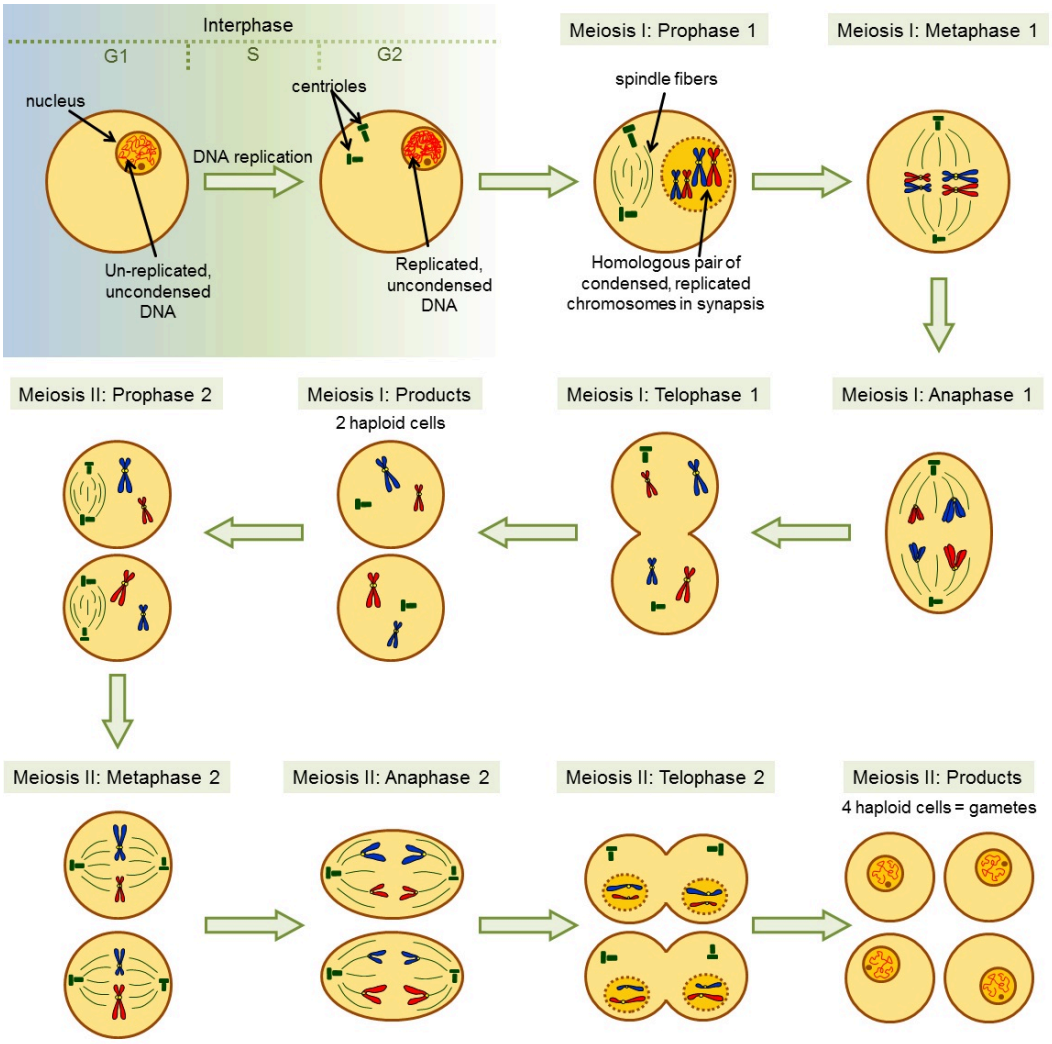


Figure 14.4 This image illustrates meiosis for a cell with 2 pairs of homologous chromosomes. The homologous pairs are based on size and the color refers to the maternal and paternal chromosomes. Work by Eva Horne.

Interphase

Meiosis is preceded by an interphase consisting of the G_1, S, and G_2 phases, which are nearly identical to the phases preceding mitosis. The G_1 phase is the first phase of interphase and is focused on cell growth. In the S phase, the DNA of the chromosomes is replicated. Finally, in the G_2 phase, the cell undergoes the final preparations for meiosis.

During DNA duplication of the S phase, each chromosome becomes composed of two identical copies (called sister chromatids) that are held together at the centromere until they are pulled apart during meiosis II. In an animal cell, the centrosomes that organize the microtubules of the meiotic spindle also replicate. This prepares the cell for the first meiotic phase.

Meiosis I

In prophase I, the chromosomes can be seen clearly microscopically. As the nuclear envelope begins to break down, the proteins associated with homologous chromosomes bring the pair close to each other. The tight pairing of the homologous chromosomes is called synapsis. In synapsis, the genes on the chromatids of the homologous chromosomes are precisely aligned with each other. An exchange of chromosome segments between non-sister homologous chromatids occurs and is called **crossing over**.

The crossover events are the first source of genetic variation produced by meiosis. A single crossover event between homologous non-sister chromatids leads to a reciprocal exchange of equivalent DNA between a maternal chromosome and a paternal chromosome. Now, when that sister chromatid is moved into a gamete, it will carry some DNA from one parent of the individual and some DNA from the other parent. The recombinant sister chromatid has a combination of maternal

and paternal genes that did not exist before the crossover. In the next class period, we will explore how crossing over and recombinant sister chromatids influence patterns of inheritance.

The key event in metaphase I is the attachment of the spindle fiber microtubules to the kinetochore proteins at the centromeres. The microtubules assembled from centrosomes at opposite poles of the cell grow toward the middle of the cell. The homologous chromosomes are held together and form a tetrad. At the end of metaphase I, each tetrad is attached to microtubules from both poles, with one homologous chromosome attached at one pole and the other homologous chromosome attached to the other pole. In addition, the nuclear membrane has broken down entirely.

By the end metaphase I, the homologous chromosomes are arranged in the center of the cell with the kinetochores facing opposite poles. The orientation of each pair of homologous chromosomes at the center of the cell is random.

This randomness, called **independent assortment**, is the physical basis for the generation of the second form of genetic variation in offspring. Consider that the homologous chromosomes of a sexually reproducing organism are originally inherited as two separate sets, one from each parent. Using humans as an example, one set of 23 chromosomes is present in the egg donated by the mother. The father provides the other set of 23 chromosomes in the sperm that fertilizes the egg. In metaphase I, these pairs line up at the midway point between the two poles of the cell. Because there is an equal chance that a microtubule fiber will encounter a maternally or paternally inherited chromosome, the arrangement of the tetrads at the metaphase plate is random. Any maternally inherited chromosome may face either pole. Any paternally inherited chromosome may also face either pole. The orientation of each tetrad is independent of the orientation of the other 22 tetrads.

In each cell that undergoes meiosis, the arrangement of the tetrads is different. The number of variations depends on the number of chromosomes making up a set. There are two possibilities for orientation (for each tetrad); thus, the possible number of alignments equals 2^n where n is the number of chromosomes per set. Humans have 23 chromosome pairs, which results in over eight million (2^{23}) possibilities. This number does not include the variability previously created in the sister chromatids by crossover. Given these two mechanisms, it is highly unlikely that any two haploid cells resulting from meiosis will have the same genetic composition (**Figure 14.5**).

To summarize the genetic consequences of meiosis I: the maternal and paternal genes are recombined by crossover events occurring on each homologous pair during prophase I; in addition, the random assortment of tetrads at metaphase produces a unique combination of maternal and paternal chromosomes that will make their way into the gametes.

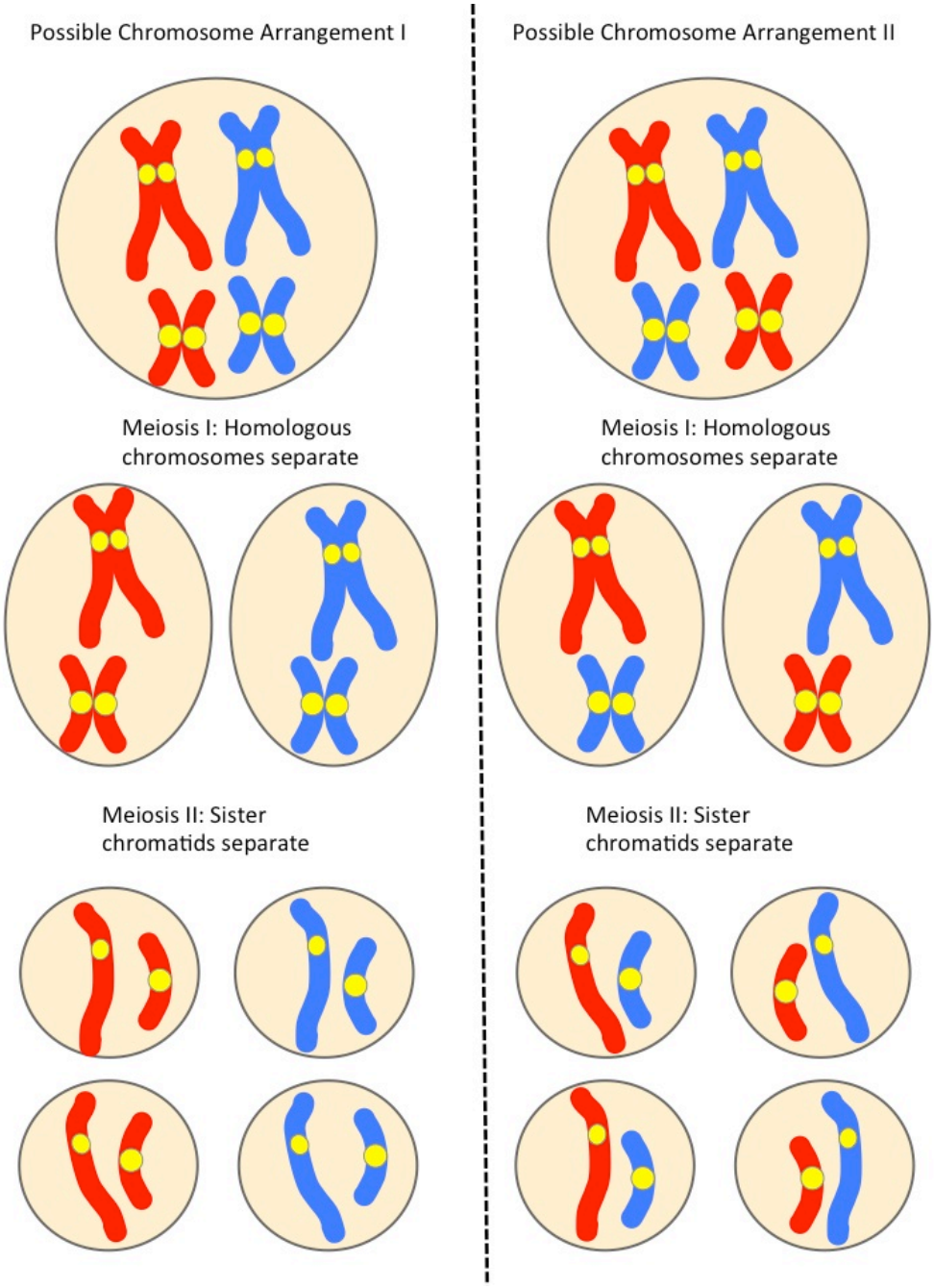


Figure 14.5 This illustration shows that, in a cell with 2 pairs of homologous chromosomes, 2 possible arrangements of chromosomes in Meiosis I will give rise to 4 different kinds of gametes. These are shown at the bottom of the figure; although there are 8 total gametes, there are 4 pairs that are identical.

In anaphase I, the spindle fibers pull the homologous chromosomes apart. The sister chromatids remain tightly bound together at the centromere (**Figure 14.4**).

In telophase I, the separated chromosomes arrive at opposite poles. The remainder of the typical telophase events may or may not occur depending on the species. In some organisms, the chromosomes decondense and nuclear envelopes form around the chromatids in telophase I.

Cytokinesis, the physical separation of the cytoplasmic components into two daughter cells, occurs without reformation of the nuclei in other organisms. In nearly all species, cytokinesis separates the cell contents by either a cleavage furrow (in animals and some fungi), or a cell plate that will ultimately lead to formation of cell walls that separate the two daughter cells (in plants). At each pole, there is just one member of each pair of the homologous chromosomes, so only one full set of the chromosomes is present. This is why the cells are considered haploid—there is only one chromosome set, even though

there are duplicate copies of the set because each homolog still consists of two sister chromatids that are still attached to each other. However, although the sister chromatids were once duplicates of the same chromosome, they are no longer identical at this stage because of crossovers.

Meiosis II

In meiosis II, the connected sister chromatids remaining in the haploid cells from meiosis I will be split to form four haploid cells (**Figure 14.4**). In some species, cells enter a brief interphase, or interkinesis, that lacks an S phase, before entering meiosis II. Chromosomes are not duplicated during interkinesis. The two cells produced in meiosis I go through the events of meiosis II in synchrony. Overall, meiosis II resembles the mitotic division of a haploid cell.

In prophase II, if the chromosomes decondensed in telophase I, they condense again. If nuclear envelopes were formed, they fragment into vesicles. The centrosomes duplicated during interkinesis move away from each other toward opposite poles, and new spindles are formed. In early metaphase II, the nuclear envelopes are completely broken down, and the spindle is fully formed. Each sister chromatid forms an individual kinetochore that attaches to microtubules from opposite poles. In metaphase II, the sister chromatids are maximally condensed and aligned at the center of the cell. In anaphase II, the sister chromatids are pulled apart by the spindle fibers and move toward opposite poles.

In telophase II, the chromosomes arrive at opposite poles and begin to decondense. Nuclear envelopes form around the chromosomes. Cytokinesis separates the two cells into four genetically unique haploid cells. At this point, the nuclei in the newly produced cells are both haploid and have only one copy of the single set of chromosomes. The cells produced are genetically unique because of the random assortment of paternal and maternal homologs and because of the recombination of maternal and paternal segments of chromosomes—with their sets of genes—that occurs during crossover.

14.3 | Mendel’s Experiments

Introduction

> “"Mendel's forty-odd year stint at the Brno abbey was, indeed, deeply constrained by rules, habits, and limits. He began his experiments on inheritance by breeding field mice but was asked to discontinue them because forcing mice to mate was considered too risqué for a monk."”

S. Mukherjee, "On Tenderness", introduction to The Best American Science and Nature Writing, Mariner Press, Boston and New York, 2013

Figure 14.6 Johann Gregor Mendel set the framework for the study of genetics.

Johann Gregor Mendel (1822–1884) (**Figure 14.6**) was a lifelong learner, teacher, scientist, and man of faith. As a young adult, he joined the Augustinian Abbey of St. Thomas in Brno in what is now the Czech Republic. Supported by the monastery, he taught physics, botany, and natural science courses at the secondary and university levels. In 1856, he began a decade-long research pursuit involving inheritance patterns in honeybees and plants, ultimately settling on pea plants as his primary model system (a system with convenient characteristics that is used to study a specific biological phenomenon to gain understanding to be applied to other systems). In 1865, Mendel presented the results of his experiments with nearly 30,000 pea plants to the local natural history society. He demonstrated that traits are transmitted faithfully from parents to offspring in specific patterns. In 1866, he published his work, *Experiments in Plant Hybridization*,[2] in the proceedings of the Natural History Society of Brünn.

Mendel's work went virtually unnoticed by the scientific community, which incorrectly believed that the process of inheritance involved a blending of parental traits that produced an intermediate physical appearance in offspring. This hypothetical process appeared to be correct because of what we know now as continuous variation. Continuous variation is the range of small differences we see among individuals in a characteristic like human height. It does appear that offspring are a "blend" of their parents' traits when we look at characteristics that exhibit continuous variation. Mendel worked instead with traits that show discontinuous variation. Discontinuous variation is the variation seen among individuals when each individual shows one of two—or a very few—easily distinguishable traits, such as violet or white flowers. Mendel's choice of these kinds of traits allowed him to see experimentally that the traits were not blended in the offspring as would have been expected at the time, but that they were inherited as distinct traits. In 1868, Mendel became abbot of the monastery and exchanged his scientific pursuits for his pastoral duties. He was not recognized for his extraordinary scientific contributions during his lifetime; in fact, it was not until 1900 that his work was rediscovered, reproduced, and revitalized by scientists on the brink of discovering the chromosomal basis of heredity.

Mendel's Crosses

Mendel's seminal work was accomplished using the garden pea, *Pisum sativum*, to study inheritance. This species naturally self-fertilizes, meaning that pollen encounters ova within the same flower. The flower petals remain sealed tightly until pollination is completed to prevent the pollination of other plants. The result is highly inbred, or "true-breeding," pea plants. These are plants that always produce offspring that look like the parent. By experimenting with true-breeding pea plants, Mendel avoided the appearance of unexpected traits in offspring that might occur if the plants were not true breeding. The garden pea also grows to maturity within one season, meaning that several generations could be evaluated over a relatively short time. Finally, large quantities of garden peas could be cultivated simultaneously, allowing Mendel to conclude that his results did not come about simply by chance.

Mendel performed hybridizations, which involve mating two true-breeding individuals that have different traits. In the pea, which is naturally self-pollinating, this is done by manually transferring pollen from the anther of a mature pea plant of one variety to the stigma of a separate mature pea plant of the second variety.

Plants used in first-generation crosses were called **P**, or parental generation, plants (**Figure 14.7**). Mendel collected the

2. Johann Gregor Mendel, "Versuche über Pflanzenhybriden." *Verhandlungen des naturforschenden Vereines in Brünn*, Bd. IV für das Jahr, 1865 Abhandlungen (1866):3–47. [for English translation, see http://www.mendelweb.org/Mendel.plain.html]

seeds produced by the P plants that resulted from each cross and grew them the following season. These offspring were called the **F_1**, or the first filial (filial = daughter or son), generation. Once Mendel examined the characteristics in the F_1 generation of plants, he allowed them to self-fertilize naturally. He then collected and grew the seeds from the F_1 plants to produce the **F_2**, or second filial, generation. Mendel's experiments extended beyond the F_2 generation to the F_3 generation, F_4 generation, and so on, but it was the ratio of characteristics in the P, F_1, and F_2 generations that were the most intriguing and became the basis of Mendel's postulates.

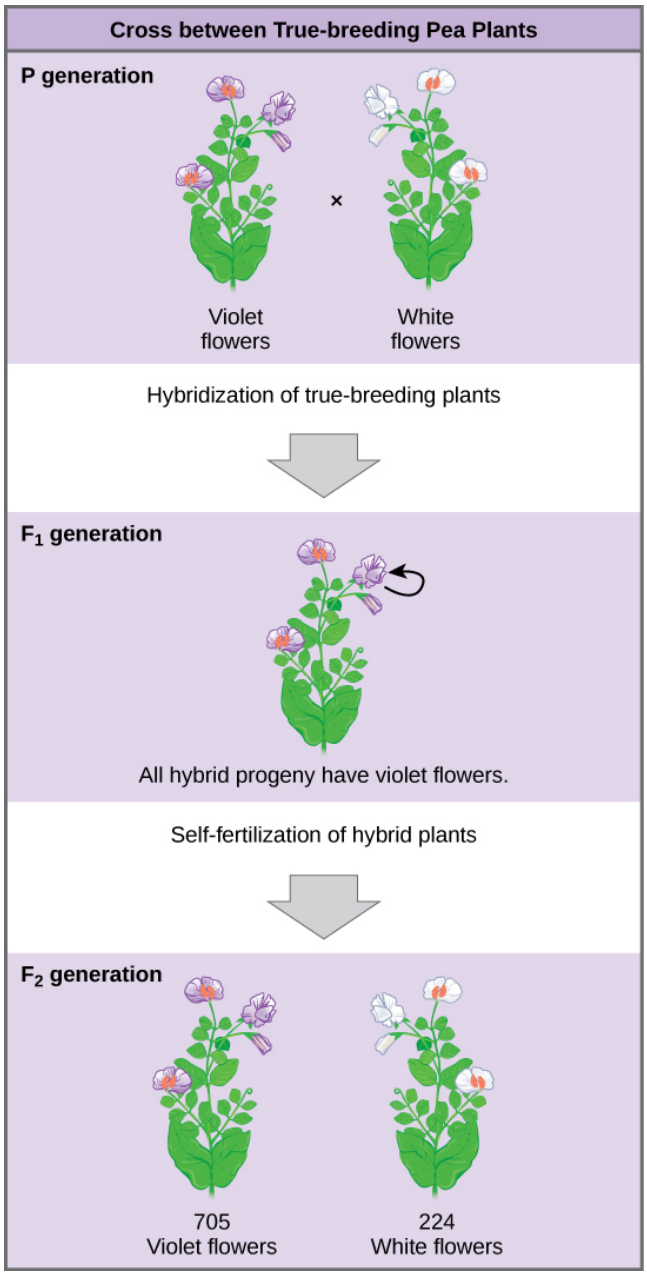


Figure 14.7 Mendel's process for performing crosses included examining flower color.

Garden Pea Characteristics Revealed the Basics of Heredity

In his 1865 publication, Mendel reported the results of his crosses involving seven different characteristics, each with two contrasting traits. A **trait** is defined as a variation in the physical appearance of a heritable characteristic. The characteristics included plant height, seed texture, seed color, flower color, pea-pod size, pea-pod color, and flower position. For the characteristic of flower color, for example, the two contrasting traits were white versus violet. To fully examine each characteristic, Mendel generated large numbers of F_1 and F_2 plants and reported results from thousands of F_2 plants.

What results did Mendel find in his crosses for flower color? First, Mendel confirmed that he was using plants that bred true

for white or violet flower color. Irrespective of the number of generations that Mendel examined, all self-crossed offspring of parents with white flowers had white flowers, and all self-crossed offspring of parents with violet flowers had violet flowers. In addition, Mendel confirmed that, other than flower color, the pea plants were physically identical. This was an important check to make sure that the two varieties of pea plants only differed with respect to one trait, flower color.

Once these validations were complete, Mendel applied the pollen from a plant with violet flowers to the stigma of a plant with white flowers. After gathering and sowing the seeds that resulted from this cross, Mendel found that 100 percent of the F_1 hybrid generation had violet flowers. Conventional wisdom at that time would have predicted the hybrid flowers to be pale violet or for hybrid plants to have equal numbers of white and violet flowers. In other words, the contrasting parental traits were expected to blend in the offspring. Instead, Mendel's results demonstrated that the white flower trait had completely disappeared in the F_1 generation.

Importantly, Mendel did not stop his experimentation there. He allowed the F_1 plants to self-fertilize and found that 705 plants in the F_2 generation had violet flowers and 224 had white flowers. This was a ratio of 3.15 violet flowers to one white flower, or approximately 3:1. When Mendel transferred pollen from a plant with violet flowers to the stigma of a plant with white flowers and vice versa, he obtained approximately the same ratio irrespective of which parent—male or female—contributed which trait. This is called a reciprocal cross —a paired cross in which the respective traits of the male and female in one cross become the respective traits of the female and male in the other cross. For the other six characteristics that Mendel examined, the F_1 and F_2 generations behaved in the same way that they behaved for flower color. One of the two traits would disappear completely from the F_1 generation, only to reappear in the F_2 generation at a ratio of roughly 3:1 (**Figure 14.8**). He wrote, in the typically sparse prose of a scientist, "In this generation, along with the dominating traits, the recessive ones also reappear, their individuality fully revealed, and they do so in the decisively expressed average proportion of 3:1, so that among each four plants of this generation three receive the dominating and one the recessive characteristic." (Gregor Mendel, '*Experiments on Plant Hybrids*', 1865)

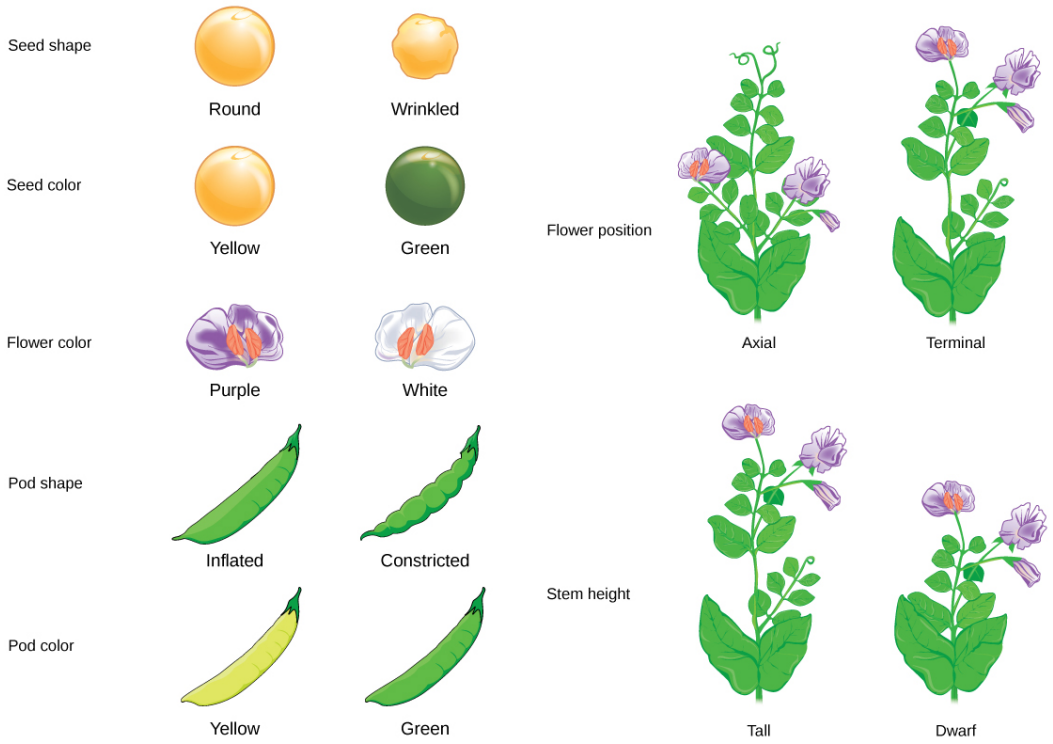


Figure 14.8 Mendel identified seven pea plant characteristics.

Upon compiling his results for many thousands of plants, Mendel concluded that the characteristics could be divided into expressed and latent traits. He called these dominant and recessive traits, respectively. **Dominant** traits are those that are inherited unchanged in a hybridization. **Recessive** traits become latent, or disappear in the offspring of a hybridization. The recessive trait does, however, reappear in the progeny of the hybrid offspring. An example of a dominant trait is the violet-colored flower trait. For this same characteristic (flower color), white-colored flowers are a recessive trait. The fact that the recessive trait reappeared in the F_2 generation meant that the traits remained separate (and were not blended) in the plants of the F_1 generation. Mendel proposed that this was because the plants possessed two copies of the trait for the flower-color characteristic, and that each parent transmitted one of their two copies to their offspring, where they came together.

Moreover, the physical observation of a dominant trait could mean that the genetic composition of the organism included two dominant versions of the characteristic, or that it included one dominant and one recessive version. Conversely, the observation of a recessive trait meant that the organism lacked any dominant versions of this characteristic.

14.4 | Laws of Inheritance

Introduction

> "If A denotes one of the two constant traits, for example, the dominating one, a the recessive, and the Aa the hybrid form in which both are united, then the expression: A + 2Aa + a gives the series for the progeny of plants hybrid in a pair of differing traits."

Gregor Mendel, 1865

The seven characteristics that Mendel evaluated in his pea plants were each expressed as one of two versions, or traits. Mendel deduced from his results that each individual had two discrete copies of the characteristic that are passed individually to offspring. We now call those two copies genes, which are carried on chromosomes. The reason we have two copies of each gene is that we inherit one from each parent. In fact, it is the chromosomes we inherit and the two copies of each gene are located on paired chromosomes. Recall that in meiosis these chromosomes are separated out into haploid gametes. This separation, or segregation, of the homologous chromosomes means also that only one of the copies of the gene gets moved into a gamete. The offspring are formed when that gamete unites with one from another parent and the two copies of each gene (and chromosome) are restored.

For cases in which a single gene controls a single characteristic, a diploid organism has two genetic copies that may or may not encode the same version of that characteristic. For example, one individual may carry a gene that determines white flower color and a gene that determines violet flower color. **Gene variants** that arise by mutation and exist at the same relative locations on homologous chromosomes are called **alleles**. Mendel examined the inheritance of genes with just two allele forms, but it is common to encounter more than two alleles for any given gene in a natural population.

Phenotypes and Genotypes

Two alleles for a given gene in a diploid organism are expressed and interact to produce physical characteristics. The observable traits expressed by an organism are referred to as its **phenotype**. An organism's underlying genetic makeup, consisting of both the physically visible and the non-expressed alleles, is called its **genotype**. Mendel's hybridization experiments demonstrate the difference between phenotype and genotype. For example, the phenotypes that Mendel observed in his crosses between pea plants with differing traits are connected to the diploid genotypes of the plants in the P, F_1, and F_2 generations. We will use a second trait that Mendel investigated, seed color, as an example. Seed color is governed by a single gene with two alleles. The yellow-seed allele is dominant and the green-seed allele is recessive. When true-breeding plants were cross-fertilized, in which one parent had yellow seeds and one had green seeds, all of the F_1 hybrid offspring had yellow seeds. That is, the hybrid offspring were phenotypically identical to the true-breeding parent with yellow seeds. However, we know that the allele donated by the parent with green seeds was not simply lost because it reappeared in some of the F_2 offspring (**Figure 14.9**). Therefore, the F_1 plants must have been genotypically different from the parent with yellow seeds.

The P plants that Mendel used in his experiments were each homozygous for the trait he was studying. Diploid organisms that are **homozygous** for a gene have two identical alleles, one on each of their homologous chromosomes. The genotype is often written as *YY* or *yy*, for which each letter represents one of the two alleles in the genotype. The dominant allele is capitalized and the recessive allele is lower case. The letter used for the gene (seed color in this case) is usually related to the dominant trait (yellow allele, in this case, or "*Y*"). Mendel's parental pea plants always bred true because both produced gametes carried the same allele. When P plants with contrasting traits were cross-fertilized, all of the offspring were **heterozygous** for the contrasting trait, meaning their genotype had different alleles for the gene being examined. For example, the F_1 yellow plants that received a *Y* allele from their yellow parent and a *y* allele from their green parent had the genotype *Yy*.

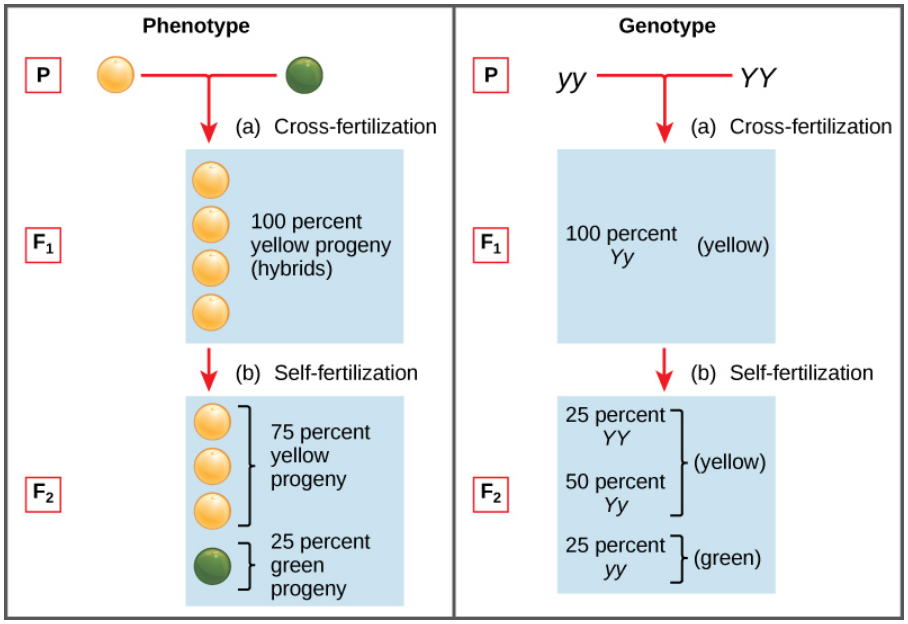


Figure 14.9 Phenotypes are physical expressions of traits that are transmitted by alleles. Capital letters represent dominant alleles and lowercase letters represent recessive alleles. The phenotypic ratios are the ratios of visible characteristics. The genotypic ratios are the ratios of gene combinations in the offspring, and these are not always distinguishable in the phenotypes.

Law of Dominance

Our discussion of homozygous and heterozygous organisms brings us to why the F_1 heterozygous offspring were identical to one of the parents, rather than expressing both alleles. In all seven pea-plant characteristics, one of the two contrasting alleles was dominant, and the other was recessive. Mendel called the dominant allele the expressed unit factor; the recessive allele was referred to as the latent unit factor. We now know that these so-called unit factors are actually genes on homologous chromosomes. For a gene that is expressed in a dominant and recessive pattern, homozygous dominant and heterozygous organisms will look identical (that is, they will have different genotypes but the same phenotype), and the recessive allele will only be observed in homozygous recessive individuals (Table 14.1).

Correspondence between Genotype and Phenotype for a Dominant-Recessive Characteristic.

	Homozygous	Heterozygous	Homozygous
Genotype	*YY*	*Yy*	*yy*
Phenotype	yellow	yellow	green

Table 14.1

Mendel's **law of dominance** states that in a heterozygote, one trait will conceal the presence of another trait for the same characteristic. For example, when crossing true-breeding violet-flowered plants with true-breeding white-flowered plants, all of the offspring were violet-flowered, even though they all had one allele for violet and one allele for white. Rather than both alleles contributing to a phenotype, the dominant allele will be expressed exclusively. The recessive allele will remain latent, but will be transmitted to offspring in the same manner as that by which the dominant allele is transmitted. The recessive trait will only be expressed by offspring that have two copies of this allele (Figure 14.10), and these offspring will breed true when self-crossed.

Figure 14.10 The allele for albinism, expressed here in humans, is recessive. Both of this child's parents carried the recessive allele.

Monohybrid Cross and the Punnett Square

When fertilization occurs between two true-breeding parents that differ by only the characteristic being studied, the process is called a **monohybrid** cross, and the resulting offspring are called monohybrids. Mendel performed seven types of monohybrid crosses, each involving contrasting traits for different characteristics. Out of these crosses, all of the F_1 offspring had the phenotype of one parent, and the F_2 offspring had a 3:1 phenotypic ratio. On the basis of these results, Mendel postulated that each parent in the monohybrid cross contributed one of two paired unit factors to each offspring, and every possible combination of unit factors was equally likely.

The results of Mendel's research can be explained in terms of probabilities, which are mathematical measures of likelihood. The probability of an event is calculated by the number of times the event occurs divided by the total number of opportunities for the event to occur. A probability of one (100 percent) for some event indicates that it is guaranteed to occur, whereas a probability of zero (0 percent) indicates that it is guaranteed to not occur, and a probability of 0.5 (50 percent) means it has an equal chance of occurring or not occurring.

To demonstrate this with a monohybrid cross, consider the case of true-breeding pea plants with yellow versus green seeds. The dominant seed color is yellow; therefore, the parental genotypes were *YY* for the plants with yellow seeds and *yy* for the plants with green seeds. A **Punnett square**, devised by the British geneticist Reginald Punnett, is useful for determining probabilities because it is drawn to predict all possible outcomes of all possible random fertilization events and their expected frequencies. **Figure 14.13** shows a Punnett square for a cross between a plant with yellow peas and one with green peas. To prepare a Punnett square, all possible combinations of the parental alleles (the genotypes of the gametes) are listed along the top (for one parent) and side (for the other parent) of a grid. The combinations of egg and sperm gametes are then made in the boxes in the table on the basis of which alleles are combining. Each box then represents the diploid genotype of a zygote, or fertilized egg. Because each possibility is equally likely, genotypic ratios can be determined from a Punnett square. If the pattern of inheritance (dominant and recessive) is known, the phenotypic ratios can be inferred as well. For a monohybrid cross of two true-breeding parents, each parent contributes one type of allele. In this case, only one genotype is possible in the F_1 offspring. All offspring are *Yy* and have yellow seeds.

When the F_1 offspring are crossed with each other, each has an equal probability of contributing either a *Y* or a *y* to the F_2 offspring. The result is a 1 in 4 (25 percent) probability of both parents contributing a *Y*, resulting in an offspring with a yellow phenotype; a 25 percent probability of parent A contributing a *Y* and parent B a *y*, resulting in offspring with a yellow phenotype; a 25 percent probability of parent A contributing a *y* and parent B a *Y*, also resulting in a yellow phenotype; and a (25 percent) probability of both parents contributing a *y*, resulting in a green phenotype. When counting all four possible outcomes, there is a 3 in 4 probability of offspring having the yellow phenotype and a 1 in 4 probability of offspring having the green phenotype. This explains why the results of Mendel's F_2 generation occurred in a 3:1 phenotypic ratio. Using

large numbers of crosses, Mendel was able to calculate probabilities, found that they fit the model of inheritance, and use these to predict the outcomes of other crosses.

Law of Segregation

Observing that true-breeding pea plants with contrasting traits gave rise to F_1 generations that all expressed the dominant trait and F_2 generations that expressed the dominant and recessive traits in a 3:1 ratio, Mendel proposed the **law of segregation**. This law states that paired unit factors (genes) must segregate equally into gametes such that offspring have an equal likelihood of inheriting either factor. For the F_2 generation of a monohybrid cross, the following three possible combinations of genotypes result: homozygous dominant, heterozygous, or homozygous recessive. Because heterozygotes could arise from two different pathways (receiving one dominant and one recessive allele from either parent), and because heterozygotes and homozygous dominant individuals are phenotypically identical, the law supports Mendel's observed 3:1 phenotypic ratio. The equal segregation of alleles is the reason we can apply the Punnett square to accurately predict the offspring of parents with known genotypes. The physical basis of Mendel's law of segregation is the first division of meiosis in which the homologous chromosomes with their different versions of each gene are segregated into daughter nuclei. This process was not understood by the scientific community during Mendel's lifetime (Figure 14.11).

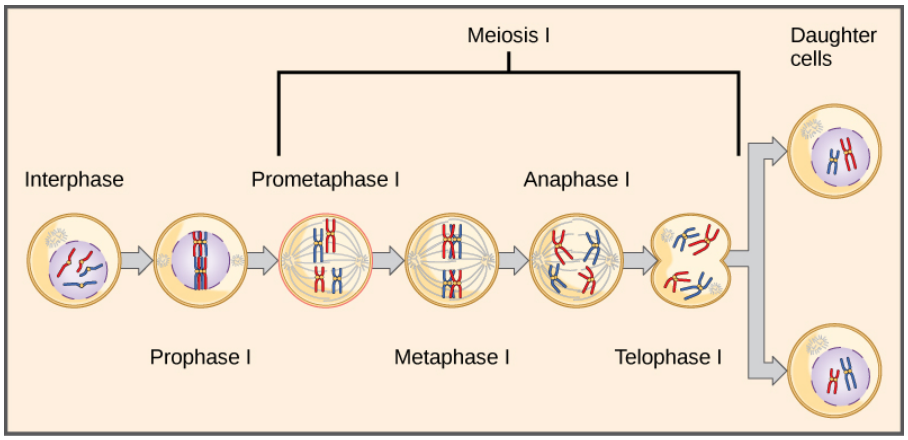


Figure 14.11 The first division in meiosis is shown.

Test Cross

Beyond predicting the offspring of a cross between known homozygous or heterozygous parents, Mendel also developed a way to determine whether an organism that expressed a dominant trait was a heterozygote or a homozygote. Called the **test cross**, this technique is still used by plant and animal breeders. In a test cross, the dominant-expressing organism is crossed with an organism that is homozygous recessive for the same characteristic. If the dominant-expressing organism is a homozygote, then all F_1 offspring will be heterozygotes expressing the dominant trait (Figure 14.12). Alternatively, if the dominant-expressing organism is a heterozygote, the F_1 offspring will exhibit a 1:1 ratio of heterozygotes and recessive homozygotes (Figure 14.12). The test cross further validates Mendel's postulate that pairs of unit factors segregate equally.

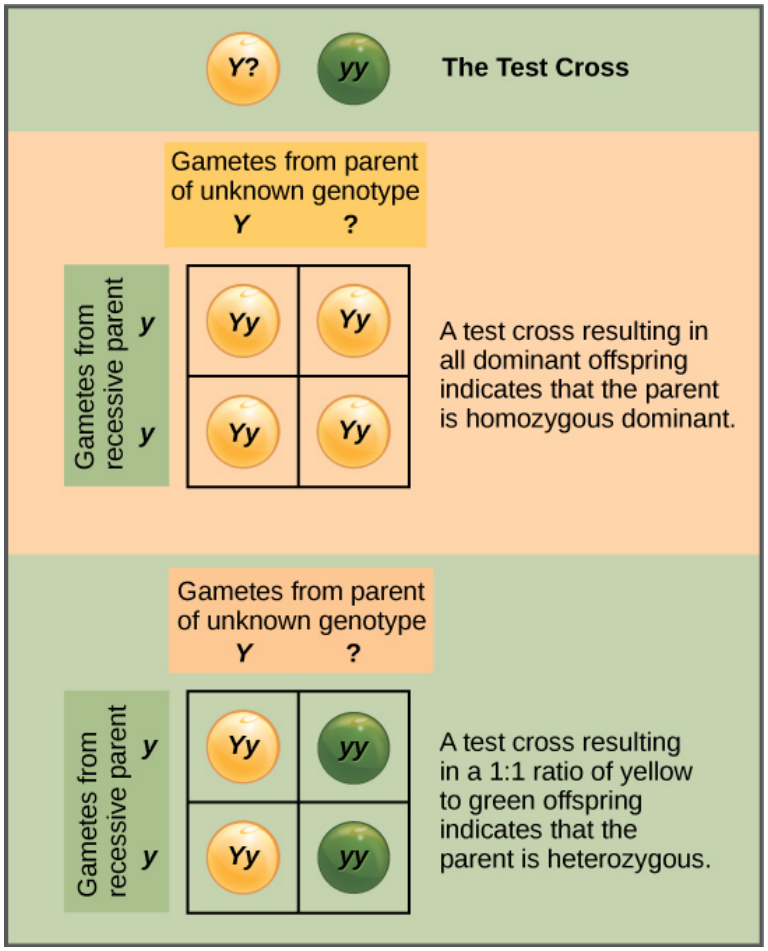


Figure 14.12 A test cross can be performed to determine whether an organism expressing a dominant trait is a homozygote or a heterozygote.

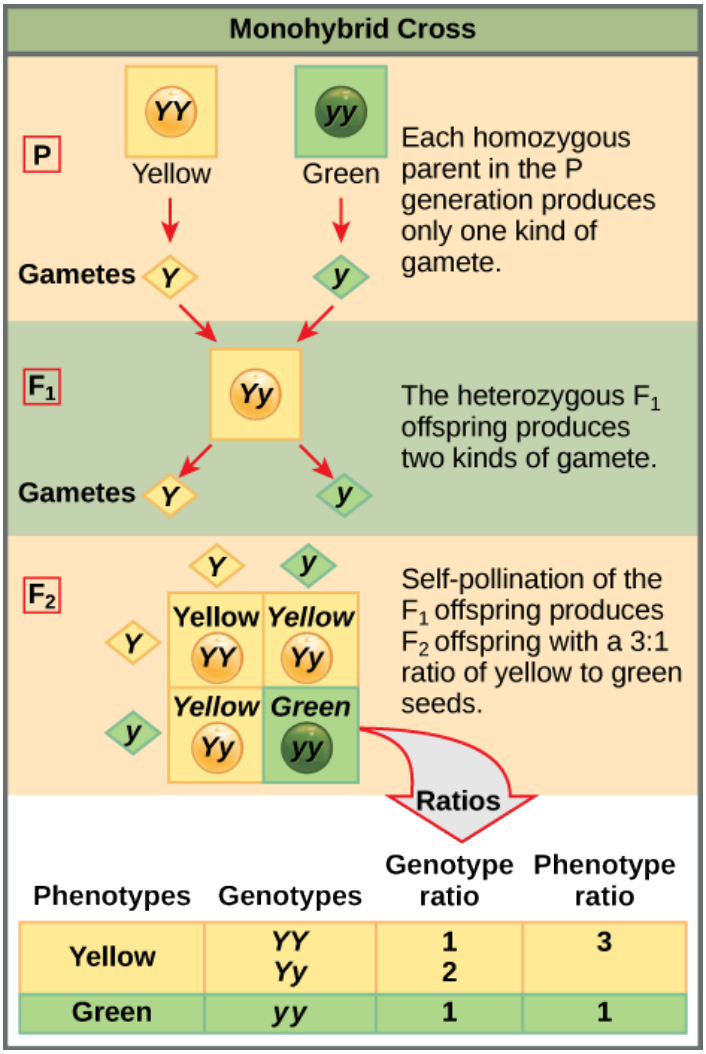


Figure 14.13 This Punnett square shows the cross between plants with yellow seeds and green seeds. The cross between the true-breeding P plants produces F_1 heterozygotes that can be self-fertilized. The self-cross of the F_1 generation can be analyzed with a Punnett square to predict the genotypes of the F_2 generation. Given an inheritance pattern of dominant–recessive, the genotypic and phenotypic ratios can then be determined.

Law of Independent Assortment

Mendel's **law of independent assortment** states that genes do not influence each other with regard to the sorting of alleles into gametes, and every possible combination of alleles for every gene is equally likely to occur. Independent assortment of genes can be illustrated by the **dihybrid** cross, a cross between two true-breeding parents that express different traits for two characteristics. Consider the characteristics of seed color and seed texture for two pea plants, one that has wrinkled, green seeds (*rryy*) and another that has round, yellow seeds (*RRYY*). Because each parent is homozygous, the law of segregation indicates that the gametes for the wrinkled–green plant all are *ry*, and the gametes for the round–yellow plant are all *RY*. Therefore, the F_1 generation of offspring all are *RrYy* (Figure 14.14).

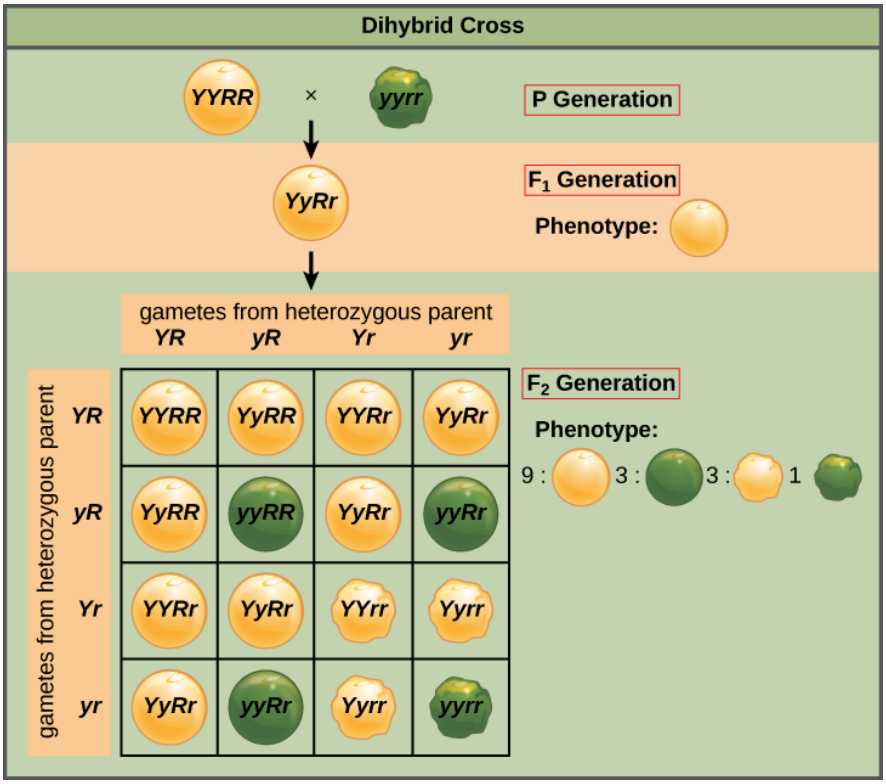


Figure 14.14 A dihybrid cross in pea plants involves the genes for seed color and texture. The P cross produces F_1 offspring that are all heterozygous for both characteristics. The resulting 9:3:3:1 F_2 phenotypic ratio is obtained using a Punnett square.

The gametes produced by the F_1 individuals must have one allele from each of the two genes. For example, a gamete could get an *R* allele for the seed shape gene and either a *Y* or a *y* allele for the seed color gene. It cannot get both an *R* and an *r* allele; each gamete can have only one allele per gene. The law of independent assortment states that a gamete into which an *r* allele is sorted would be equally likely to contain either a *Y* or a *y* allele. Thus, there are four equally likely gametes that can be formed when the *RrYy* heterozygote is self-crossed, as follows: *RY*, *rY*, *Ry*, and *ry*. Arranging these gametes along the top and left of a 4 × 4 Punnett square (**Figure 14.14**) gives us 16 equally likely genotypic combinations. From these genotypes, we find a phenotypic ratio of 9 round–yellow:3 round–green:3 wrinkled–yellow:1 wrinkled–green (**Figure 14.14**). These are the offspring ratios we would expect, assuming we performed the crosses with a large enough sample size.

The physical basis for the law of independent assortment also lies in meiosis I, in which the different homologous pairs line up in random orientations. Each gamete can contain any combination of paternal and maternal chromosomes (and therefore the genes on them) because the orientation of tetrads on the metaphase plane is random (**Figure 14.15**).

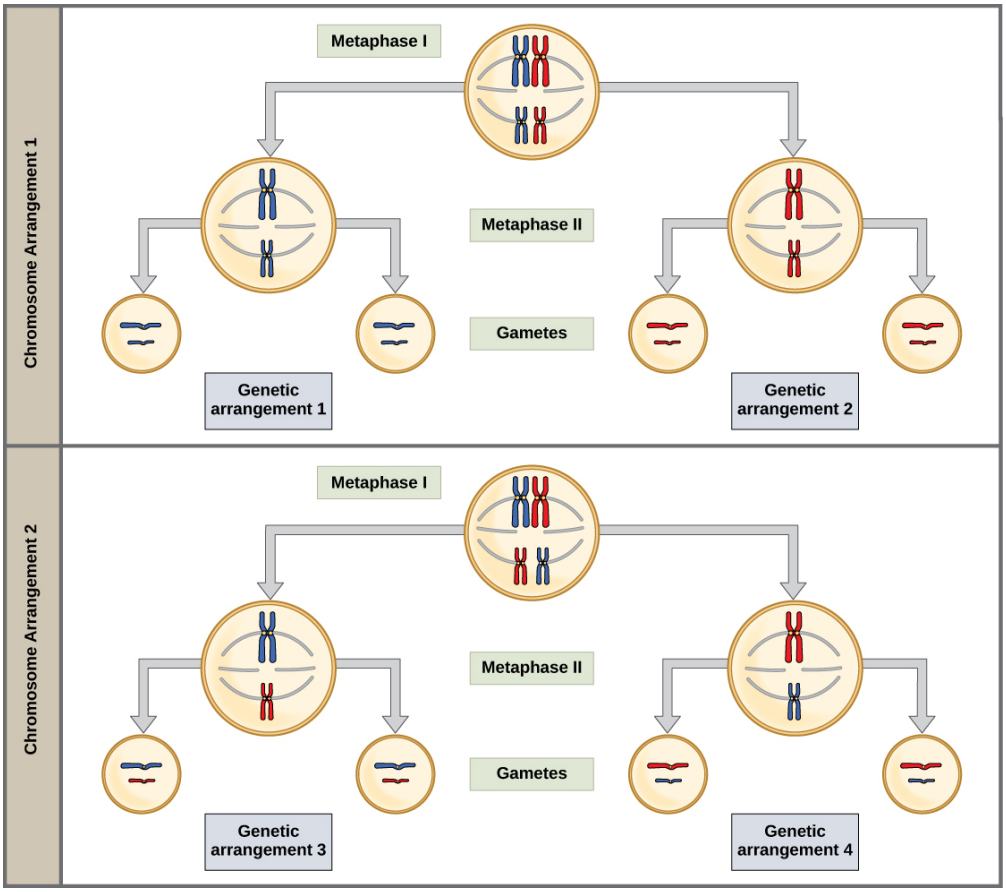


Figure 14.15 The random segregation into daughter nuclei that happens during the first division in meiosis can lead to a variety of possible genetic arrangements.

15 | VARIATIONS ON MENDELIAN GENETICS

Introduction

> "Certain students of genetics inferred that the Mendelian units responsible for the selected character were genes producing only a single effect. This was careless logic. It took a good deal of hammering to get rid of this erroneous idea. As facts accumulated it became evident that each gene produces not a single effect, but in some cases a multitude of effects on the characters of the individual. It is true that in most genetic work only one of these character-effects is selected for study—the one that is most sharply defined and separable from its contrasted character—but in most cases minor differences also are recognizable that are just as much the product of the same gene as is the major effect. "

Thomas Hunt Morgan, 1935

Mendel studied traits with only one mode of inheritance, and one very obvious phenotype, in pea plants. The inheritance of the traits he studied all followed the relatively simple pattern of dominant and recessive alleles for a single characteristic, and allowed him to suggest his Laws of Inheritance. There are several important modes of inheritance, discovered after Mendel's work, that do not follow the dominant and recessive, single-gene model. So mendel's Laws are not always going to predict the phenotypes seen after doing a genetic cross.

Alternatives to Dominance and Recessiveness

Mendel's experiments with pea plants suggested that: 1) two types of "units" or alleles exist for every gene; 2) alleles maintain their integrity in each generation (no blending); and 3) in the presence of the dominant allele, the recessive allele is hidden, with no contribution to the phenotype. Therefore, recessive alleles can be "carried" and not expressed by individuals. Such heterozygous individuals are sometimes referred to as "carriers." Since then, genetic studies in other organisms have shown that much more complexity exists, but that the fundamental principles of Mendelian genetics still hold true. In the sections to follow, we consider some of the extensions of Mendelism.

Incomplete Dominance

Mendel's results, demonstrating that traits are inherited as dominant and recessive pairs, contradicted the view at that time that offspring exhibited a blend of their parents' traits. However, the heterozygote phenotype occasionally does appear to be intermediate between the two parents. For example, in the snapdragon, *Antirrhinum majus* (**Figure 15.1**), a cross between a homozygous parent with white flowers (C^WC^W) and a homozygous parent with red flowers (C^RC^R) will produce offspring with pink flowers (C^RC^W). (Note that different genotypic abbreviations are used for Mendelian extensions to distinguish these patterns from simple dominance and recessiveness.) This pattern of inheritance is described as **incomplete dominance**, meaning that one of the alleles appears in the phenotype in the heterozygote, but not to the exclusion of the other, which can also be seen. The allele for red flowers is incompletely dominant over the allele for white flowers. However, the results of a heterozygote self-cross can still be predicted, just as with Mendelian dominant and recessive crosses. In this case, the genotypic ratio would be 1 C^RC^R:2 C^RC^W:1 C^WC^W, and the phenotypic ratio would be 1:2:1 for red:pink:white. The basis for the intermediate color in the heterozygote is simply that the pigment produced by the red allele (anthocyanin) is diluted in the heterozygote and therefore appears pink because of the white background of the flower petals.

Figure 15.1 These pink flowers of a heterozygote snapdragon result from incomplete dominance. (credit: "storebukkebruse"/Flickr)

Codominance

A variation on incomplete dominance is **codominance**, in which both alleles for the same characteristic are simultaneously expressed in the heterozygote. An example of codominance occurs in the ABO blood groups of humans. The A and B alleles are expressed in the form of A or B molecules present on the surface of red blood cells. Homozygotes (I^AI^A and I^BI^B) express either the A or the B phenotype, and heterozygotes (I^AI^B) express both phenotypes equally. The I^AI^B individual has blood type AB. In a self-cross between heterozygotes expressing a codominant trait, the three possible offspring genotypes are phenotypically distinct. However, the 1:2:1 genotypic ratio characteristic of a Mendelian monohybrid cross still applies (**Figure 15.2**).

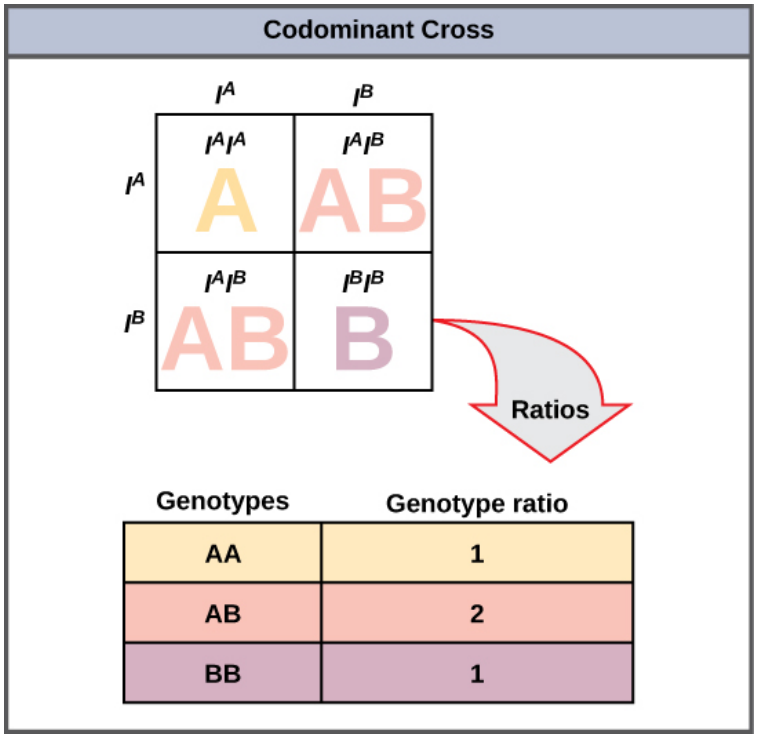


Figure 15.2 This Punnet square shows an AB/AB blood type cross

Multiple Alleles

Mendel implied that only two alleles, one dominant and one recessive, could exist for a given gene. We now know that this is an oversimplification. Although individual humans (and all diploid organisms) can only have two alleles for a given gene, multiple alleles may exist at the population level, such that many combinations of two alleles are observed. Note that when many alleles exist for the same gene, the convention is to denote the most common phenotype or genotype in the natural population as the **wild type** (often abbreviated "+"). All other phenotypes or genotypes are considered variants (mutants) of this typical form, meaning they deviate from the wild type. The variant may be recessive or dominant to the wild-type allele.

An example of multiple alleles is the ABO blood-type system in humans. In this case, there are three alleles circulating in the population. The I^A allele codes for A molecules on the red blood cells, the I^B allele codes for B molecules on the surface of red blood cells, and the *i* allele codes for no molecules on the red blood cells. In this case, the I^A and I^B alleles are codominant with each other and are both dominant over the *i* allele. Although there are three alleles present in a population, each individual only gets two of the alleles from their parents. This produces the genotypes and phenotypes shown in **Figure 15.3**. Notice that instead of three genotypes, there are six different genotypes when there are three alleles. The number of possible phenotypes depends on the dominance relationships between the three alleles.

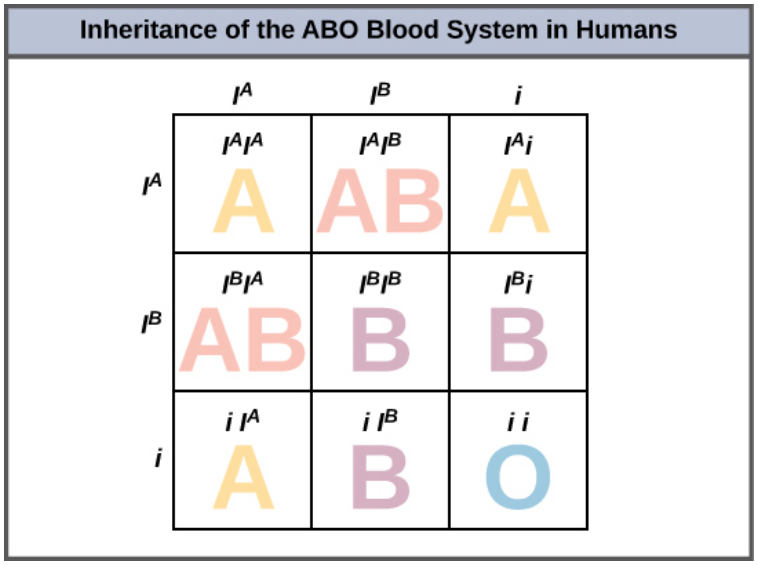


Figure 15.3 Inheritance of the ABO blood system in humans is shown.

evolution IN ACTION

Multiple Alleles Confer Drug Resistance in the Malaria Parasite

Malaria is a parasitic disease in humans that is transmitted by infected female mosquitoes, including *Anopheles gambiae*, and is characterized by cyclic high fevers, chills, flu-like symptoms, and severe anemia. *Plasmodium falciparum* and *P. vivax* are the most common causative agents of malaria, and *P. falciparum* is the most deadly. When promptly and correctly treated, *P. falciparum* malaria has a mortality rate of 0.1 percent. However, in some parts of the world, the parasite has evolved resistance to commonly used malaria treatments, so the most effective malarial treatments can vary by geographic region.

In Southeast Asia, Africa, and South America, *P. falciparum* has developed resistance to the anti-malarial drugs chloroquine, mefloquine, and sulfadoxine-pyrimethamine. *P. falciparum*, which is haploid during the life stage in which it is infective to humans, has evolved multiple drug-resistant mutant alleles of the *dhps* gene. Varying degrees of sulfadoxine resistance are associated with each of these alleles. Being haploid, *P. falciparum* needs only one drug-resistant allele to express this trait.

In Southeast Asia, different sulfadoxine-resistant alleles of the *dhps* gene are localized to different geographic regions. This is a common evolutionary phenomenon that comes about because drug-resistant mutants arise in a population and interbreed with other *P. falciparum* isolates in close proximity. Sulfadoxine-resistant parasites cause considerable human hardship in regions in which this drug is widely used as an over-the-counter malaria remedy. As is common with pathogens that multiply to large numbers within an infection cycle, *P. falciparum* evolves relatively rapidly (over a decade or so) in response to the selective pressure of commonly used anti-malarial drugs. For this reason, scientists must constantly work to develop new drugs or drug combinations to combat the worldwide malaria burden.[1]

Sex-Linked Traits

In humans, as well as in many other animals and some plants, the sex of the individual is determined by sex chromosomes—one pair of non-homologous chromosomes. Until now, we have only considered inheritance patterns among non-sex chromosomes, or autosomes. In addition to 22 homologous pairs of autosomes, human females have a homologous pair of X chromosomes, whereas human males have an XY chromosome pair. Although the Y chromosome contains a small region of similarity to the X chromosome so that they can pair during meiosis, the Y chromosome is much shorter and contains fewer genes. When a gene being examined is present on the X, but not the Y, chromosome, it is **X-linked**.

Eye color in *Drosophila*, the common fruit fly, was the first X-linked trait to be identified. Thomas Hunt Morgan mapped this trait to the X chromosome in 1910. Like humans, *Drosophila* males have an XY chromosome pair, and females are XX.

1. Sumiti Vinayak et al., "Origin and Evolution of Sulfadoxine Resistant *Plasmodium falciparum*," *PLoS Pathogens* 6 (2010): e1000830.

In flies the wild-type eye color is red (X^W) and is dominant to white eye color (X^w) (**Figure 15.4**). Because of the location of the eye-color gene, reciprocal crosses do not produce the same offspring ratios. Males are said to be hemizygous, in that they have only one allele for any X-linked characteristic. Hemizygosity makes descriptions of dominance and recessiveness irrelevant for XY males. *Drosophila* males lack the white gene on the Y chromosome; that is, their genotype can only be X^WY or X^wY. In contrast, females have two allele copies of this gene and can be X^WX^W, X^WX^w, or X^wX^w.

Figure 15.4 In *Drosophila*, the gene for eye color is located on the X chromosome. Red eye color is wild-type and is dominant to white eye color.

In an X-linked cross, the genotypes of F_1 and F_2 offspring depend on whether the recessive trait was expressed by the male or the female in the P generation. With respect to *Drosophila* eye color, when the P male expresses the white-eye phenotype and the female is homozygously red-eyed, all members of the F_1 generation exhibit red eyes (**Figure 15.5**). The F_1 females are heterozygous (X^WX^w), and the males are all X^WY, having received their X chromosome from the homozygous dominant P female and their Y chromosome from the P male. A subsequent cross between the X^WX^w female and the X^WY male would produce only red-eyed females (with X^WX^W or X^WX^w genotypes) and both red- and white-eyed males (with X^WY or X^wY genotypes). Now, consider a cross between a homozygous white-eyed female and a male with red eyes. The F_1 generation would exhibit only heterozygous red-eyed females (X^WX^w) and only white-eyed males (X^wY). Half of the F_2 females would be red-eyed (X^WX^w) and half would be white-eyed (X^wX^w). Similarly, half of the F_2 males would be red-eyed (X^WY) and half would be white-eyed (X^wY).

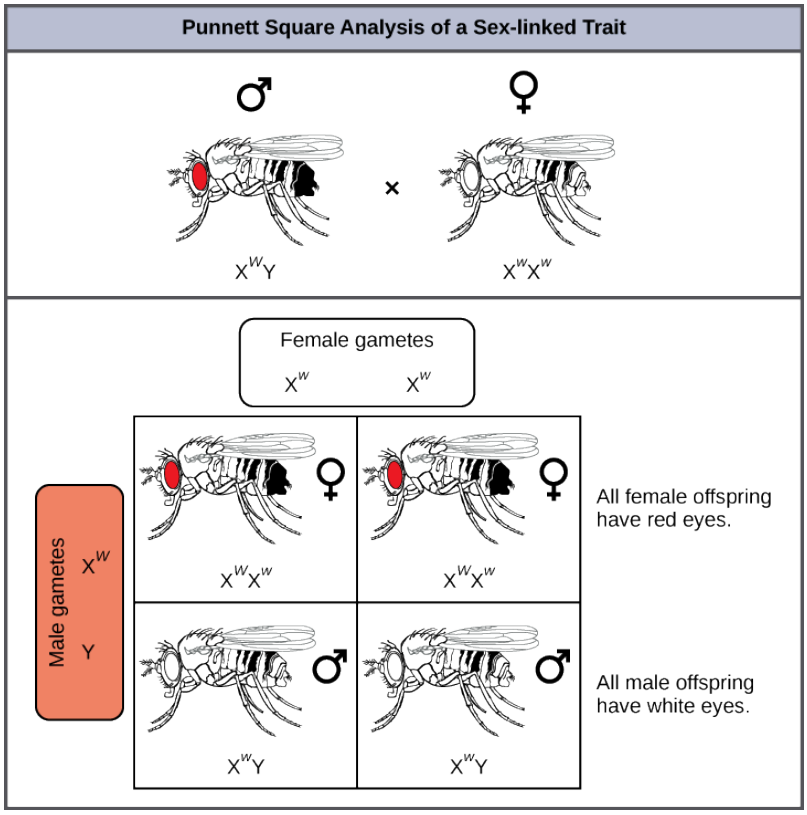


Figure 15.5 Crosses involving sex-linked traits often give rise to different phenotypes for the different sexes of offspring, as is the case for this cross involving red and white eye color in *Drosophila*. In the diagram, *w* is the white-eye mutant allele and *W* is the wild-type, red-eye allele.

Discoveries in fruit fly genetics can be applied to human genetics. When a female parent is homozygous for a recessive X-linked trait, she will pass the trait on to 100 percent of her male offspring, because the males will receive the Y chromosome from the male parent. In humans, the alleles for certain conditions (some color-blindness, hemophilia, and muscular dystrophy) are X-linked. Females who are heterozygous for these diseases are said to be carriers and may not exhibit any phenotypic effects. These females will pass the disease to half of their sons and will pass carrier status to half of their daughters; therefore, X-linked traits appear more frequently in males than females.

In some groups of organisms with sex chromosomes, the sex with the non-homologous sex chromosomes is the female rather than the male. This is the case for all birds. In this case, sex-linked traits will be more likely to appear in the female, in whom they are hemizygous.

Linked Genes Violate the Law of Independent Assortment

Although all of Mendel's pea plant characteristics behaved according to the law of independent assortment, we now know that some allele combinations are not inherited independently of each other. Genes that are located on separate, non-homologous chromosomes will always sort independently. However, each chromosome contains hundreds or thousands of genes, organized linearly on chromosomes like beads on a string. The segregation of alleles into gametes can be influenced by **linkage**, in which genes that are located physically close to each other on the same chromosome are more likely to be inherited as a pair. However, because of the process of recombination, or "crossover," it is possible for two genes on the same chromosome to behave independently, or as if they are not linked. To understand this, let us consider the biological basis of gene linkage and recombination.

Homologous chromosomes possess the same genes in the same order, though the specific alleles of the gene can be different on each of the two chromosomes. Recall that during interphase and prophase I of meiosis, homologous chromosomes first replicate and then synapse, with like genes on the homologs aligning with each other. At this stage, segments of homologous chromosomes exchange linear segments of genetic material (**Figure 15.6**). This process is called **recombination**, or crossover, and it is a common genetic process. Because the genes are aligned during recombination, the gene order is not altered. Instead, the result of recombination is that maternal and paternal alleles are combined onto the same chromosome. Across a given chromosome, several recombination events may occur, causing extensive shuffling of alleles.

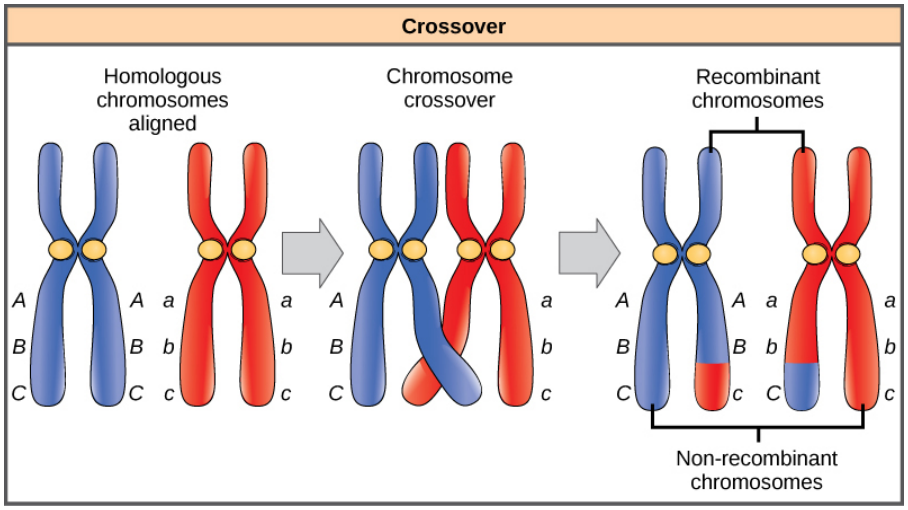


Figure 15.6 The process of crossover, or recombination, occurs when two homologous chromosomes align and exchange a segment of genetic material.

When two genes are located on the same chromosome, they are considered linked, and their alleles tend to be transmitted through meiosis together. To exemplify this, imagine a dihybrid cross involving flower color and plant height in which the genes are next to each other on the chromosome. If one homologous chromosome has alleles for tall plants and red flowers, and the other chromosome has genes for short plants and yellow flowers, then when the gametes are formed, the tall and red alleles will tend to go together into a gamete and the short and yellow alleles will go into other gametes. These are called the parental genotypes because they have been inherited intact from the parents of the individual producing gametes. But unlike if the genes were on different chromosomes, there will be no gametes with tall and yellow alleles and no gametes with short and red alleles. If you create a Punnett square with these gametes, you will see that the classical Mendelian prediction of a 9:3:3:1 outcome of a dihybrid cross would not apply. As the distance between two genes increases, the probability of one or more crossovers between them increases and the genes behave more like they are on separate chromosomes. Geneticists have used the proportion of recombinant gametes (the ones not like the parents) as a measure of how far apart genes are on a chromosome. Using this information, they have constructed linkage maps of genes on chromosomes for well-studied organisms, including humans.

Mendel's seminal publication makes no mention of linkage, and many researchers have questioned whether he encountered linkage but chose not to publish those crosses out of concern that they would invalidate his independent assortment postulate. The garden pea has seven chromosomes, and some have suggested that his choice of seven characteristics was not a coincidence. However, even if the genes he examined were not located on separate chromosomes, it is possible that he simply did not observe linkage because of the extensive shuffling effects of recombination.

16 | POPULATION GENETICS

16.1 | Population Evolution

Introduction

> “Genetics is to biology what atomic theory is to physics. Its principle is clear: that inheritance is based on particles and not on fluids. Instead of the essence of each parent mixing, with each child the blend of those who made him, information is passed on as a series of units. The bodies of successive generations transport them through time, so that a long-lost character may emerge in a distant descendant. The genes themselves may be older than the species that bear them.”

— John Stephen Jones, Almost Like a Whale: The Origin of Species Updated. Doubleday, 1999.

The mechanisms of inheritance, or genetics, were not understood at the time Charles Darwin and Alfred Russel Wallace were developing their idea of natural selection. This lack of understanding was a stumbling block to understanding many aspects of evolution. In fact, the predominant (and incorrect) genetic theory of the time, blending inheritance, made it difficult to understand how natural selection might operate. Darwin and Wallace were unaware of the genetics work by Austrian monk Gregor Mendel, which was published in 1866, not long after publication of Darwin's book, *On the Origin of Species*. Mendel’s work was rediscovered in the early twentieth century at which time geneticists were rapidly coming to an understanding of the basics of inheritance. Initially, the newly discovered particulate nature of genes made it difficult for biologists to understand how gradual evolution could occur. But over the next few decades genetics and evolution were integrated in what became known as the **modern synthesis**—the coherent understanding of the relationship between natural selection and genetics that took shape by the 1940s and is generally accepted today. In sum, the modern synthesis describes how evolutionary processes, such as natural selection, can affect a population’s genetic makeup, and, in turn, how this can result in the gradual evolution of populations and species. The theory also connects this change of a population over time, called **microevolution**, with the processes that gave rise to new species and higher taxonomic groups with widely divergent characters, called **macroevolution**.

evolution CONNECTION

Evolution and Flu Vaccines

Every fall, the media starts reporting on flu vaccinations and potential outbreaks. Scientists, health experts, and institutions determine recommendations for different parts of the population, predict optimal production and inoculation schedules, create vaccines, and set up clinics to provide inoculations. You may think of the annual flu shot as a lot of media hype, an important health protection, or just a briefly uncomfortable prick in your arm. But do you think of it in terms of evolution?

The media hype of annual flu shots is scientifically grounded in our understanding of evolution. Each year, scientists across the globe strive to predict the flu strains that they anticipate being most widespread and harmful in the coming year. This knowledge is based in how flu strains have evolved over time and over the past few flu seasons. Scientists then work to create the most effective vaccine to combat those selected strains. Hundreds of millions of doses are produced in a short period in order to provide vaccinations to key populations at the optimal time.

Because viruses, like the flu, evolve very quickly (especially in evolutionary time), this poses quite a challenge. Viruses mutate and replicate at a fast rate, so the vaccine developed to protect against last year's flu strain may not provide the protection needed against the coming year's strain. Evolution of these viruses means continued adaptions to ensure survival, including adaptations to survive previous vaccines.

Population Genetics

Recall that a gene for a particular character may have several alleles, or variants, that code for different traits associated with that character. For example, in the ABO blood type system in humans, three alleles determine the particular blood-type protein on the surface of red blood cells. Each individual in a population of diploid organisms can only carry two alleles for a particular gene, but more than two may be present in the individuals that make up the population. Mendel followed alleles as they were inherited from parent to offspring. In the early twentieth century, biologists in a field of study known as **population genetics** began to study how selective forces change a population through changes in allele and genotypic frequencies.

The **allele frequency** (or gene frequency) is the rate at which a specific allele appears within a population. Until now we have discussed evolution as a change in the characteristics of a population of organisms, but behind that phenotypic change is genetic change. In population genetics, the term evolution is defined as a change in the frequency of an allele in a population. Using the ABO blood type system as an example, the frequency of one of the alleles, I^A, is the number of copies of that allele divided by all the copies of the ABO gene in the population. For example, a study in Jordan[1] found a frequency of I^A to be 26.1 percent. The I^B and I^0 alleles made up 13.4 percent and 60.5 percent of the alleles respectively, and all of the frequencies added up to 100 percent. A change in this frequency over time would constitute evolution in the population.

The allele frequency within a given population can change depending on environmental factors; therefore, certain alleles become more widespread than others during the process of natural selection. Natural selection can alter the population's genetic makeup; for example, if a given allele confers a phenotype that allows an individual to better survive or have more offspring. Because many of those offspring will also carry the beneficial allele, and often the corresponding phenotype, they will have more offspring of their own that also carry the allele, thus, perpetuating the cycle. Over time, the allele will spread throughout the population. Some alleles will quickly become fixed in this way, meaning that every individual of the population will carry the allele, while detrimental mutations may be swiftly eliminated if derived from a dominant allele from the gene pool. The **gene pool** is the sum of all the alleles in a population.

Sometimes, allele frequencies within a population change randomly with no advantage to the population over existing allele frequencies. This phenomenon is called genetic drift. Natural selection and genetic drift usually occur simultaneously in populations and are not isolated events. It is hard to determine which process dominates because it is often nearly impossible to determine the cause of change in allele frequencies at each occurrence. An event that initiates an allele frequency change in an isolated part of the population, which is not typical of the original population, is called the **founder effect**. Natural selection, random drift, and founder effects can lead to significant changes in the genome of a population.

1. Sahar S. Hanania, Dhia S. Hassawi, and Nidal M. Irshaid, "Allele Frequency and Molecular Genotypes of ABO Blood Group System in a Jordanian Population," *Journal of Medical Sciences* 7 (2007): 51-58, doi:10.3923/jms.2007.51.58.

Hardy-Weinberg Principle of Equilibrium

In the early twentieth century, English mathematician Godfrey Hardy and German physician Wilhelm Weinberg stated the principle of equilibrium to describe the genetic makeup of a population. The theory, which later became known as the **Hardy-Weinberg principle of equilibrium**, states that a population's allele and genotype frequencies are inherently stable— unless some kind of evolutionary force is acting upon the population, neither the allele nor the genotypic frequencies would change. The **Hardy-Weinberg principle** assumes conditions with no mutations, migration, emigration, or selective pressure for or against genotype, plus an infinite population; while no population can satisfy those conditions, the principle offers a useful model against which to compare real population changes.

Working under this theory, population geneticists represent different alleles as different variables in their mathematical models. The variable p, for example, often represents the frequency of a particular allele, say Y for the trait of yellow in Mendel's peas, while the variable q represents the frequency of y alleles that confer the color green. If these are the only two possible alleles for a given locus in the population, $p + q = 1$. In other words, all the p alleles and all the q alleles make up all of the alleles for that locus that are found in the population.

But what ultimately interests most biologists is not the frequencies of different alleles, but the frequencies of the resulting genotypes, known as the population's genetic structure, from which scientists can surmise the distribution of phenotypes. If the phenotype is observed, only the genotype of the homozygous recessive alleles can be known; the calculations provide an estimate of the remaining genotypes. Since each individual carries two alleles per gene, if the allele frequencies (p and q) are known, predicting the frequencies of these genotypes is a simple mathematical calculation to determine the probability of getting these genotypes if two alleles are drawn at random from the gene pool. So in the above scenario, an individual pea plant could be pp (YY), and thus produce yellow peas; pq (Yy), also yellow; or qq (yy), and thus producing green peas (**Figure 16.1**). In other words, the frequency of pp individuals is simply p^2; the frequency of pq individuals is 2pq; and the frequency of qq individuals is q^2. And, again, if p and q are the only two possible alleles for a given trait in the population, these genotypes frequencies will sum to one: $p^2 + 2pq + q^2 = 1$.

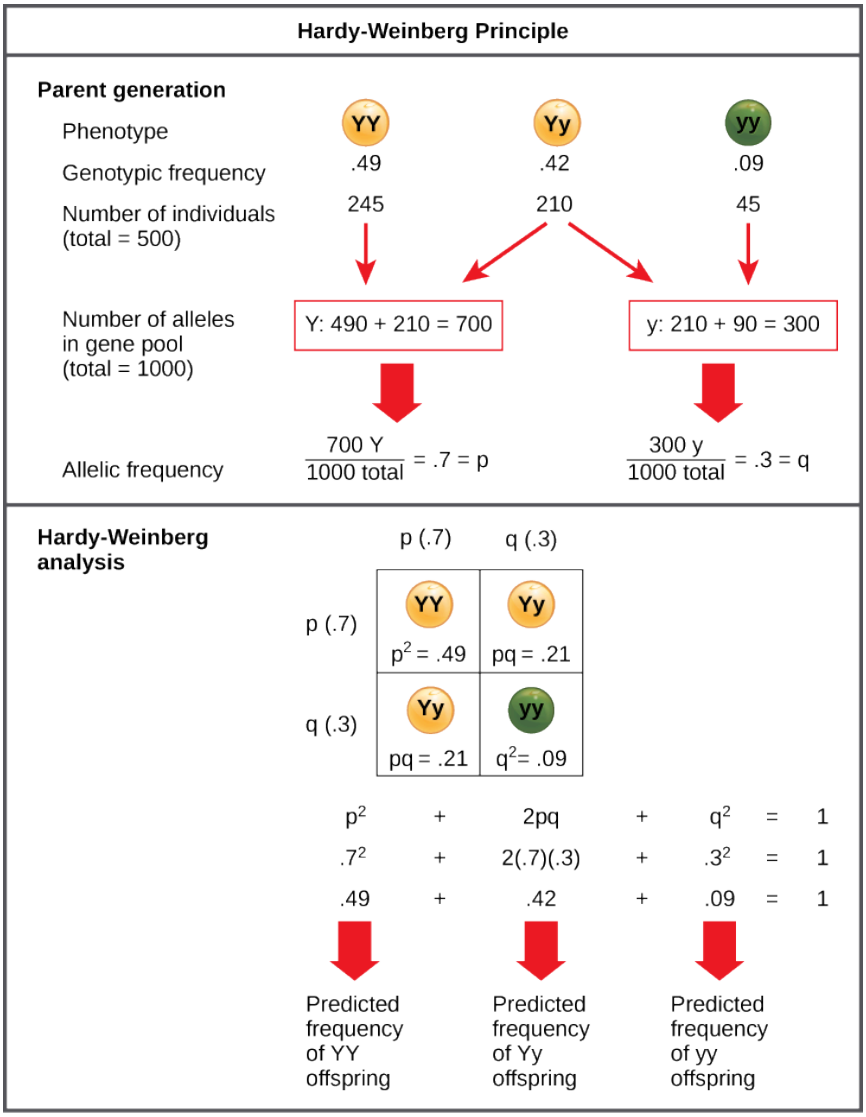


Figure 16.1 When populations are in the Hardy-Weinberg equilibrium, the allelic frequency is stable from generation to generation and the distribution of alleles can be determined from the Hardy-Weinberg equation. If the allelic frequency measured in the field differs from the predicted value, scientists can make inferences about what evolutionary forces are at play.

In theory, if a population is at equilibrium—that is, there are no evolutionary forces acting upon it—generation after generation would have the same gene pool and genetic structure, and these equations would all hold true all of the time. Of course, even Hardy and Weinberg recognized that no natural population is immune to evolution. Populations in nature are constantly changing in genetic makeup due to drift, mutation, possibly migration, and selection. As a result, the only way to determine the exact distribution of phenotypes in a population is to go out and count them. But the Hardy-Weinberg principle gives scientists a mathematical baseline of a non-evolving population to which they can compare evolving populations and thereby infer what evolutionary forces might be at play. If the frequencies of alleles or genotypes deviate from the value expected from the Hardy-Weinberg equation, then the population is evolving.

16.2 | Population Genetics

Introduction

> “ The proof given by Wright, that non-adaptive differentiation will occur in small populations owing to 'drift', or the chance fixation of some new mutation or recombination, is one of the most important results of mathematical analysis applied to the facts of neo-mendelism. It gives accident as well as adaptation a place in evolution, and at one stroke explains many facts which puzzled earlier selectionists, notably the much greater degree of divergence shown by island than mainland forms, by forms in isolated lakes than in continuous river-systems. ”

Sir Julian Huxley, 1942

The leap from understanding genes and mutations to an understanding of the evolution of populations required the identification of other mechanisms that allowed genes to become common or uncommon in populations. Individuals of a population often display different phenotypes, or express different alleles of a particular gene, referred to as polymorphisms. Populations with two or more variations of particular characteristics are called polymorphic. The distribution of phenotypes among individuals, known as the **population variation**, is influenced by a number of factors, including the population’s genetic structure and the environment (Figure 16.2). Understanding the sources of a phenotypic variation in a population is important for determining how a population will evolve in response to different evolutionary pressures.

Figure 16.2 The distribution of phenotypes in this litter of kittens illustrates population variation. (credit: Pieter Lanser)

Genetic Variance

Natural selection and some of the other evolutionary forces can only act on heritable traits, namely an organism’s genetic code. Because alleles are passed from parent to offspring, those that confer beneficial traits or behaviors may be selected for, while deleterious alleles may be selected against. Acquired traits, for the most part, are not heritable. For example, if an athlete works out in the gym every day, building up muscle strength, the athlete’s offspring will not necessarily grow up to be a body builder. If there is a genetic basis for the ability to run fast, on the other hand, this may be passed to a child.

Heritability is the fraction of phenotype variation that can be attributed to genetic differences, or genetic variance, among individuals in a population. The greater the hereditability of a population’s phenotypic variation, the more susceptible it is to the evolutionary forces that act on heritable variation.

The diversity of alleles and genotypes within a population is called genetic variance. When scientists are involved in the breeding of a species, such as with animals in zoos and nature preserves, they try to increase a population’s genetic variance to preserve as much of the phenotypic diversity as they can. This also helps reduce the risks associated with inbreeding, the mating of closely related individuals, which can have the undesirable effect of bringing together deleterious recessive mutations that can cause abnormalities and susceptibility to disease. For example, a disease that is caused by a

rare, recessive allele might exist in a population, but it will only manifest itself when an individual carries two copies of the allele. Because the allele is rare in a normal, healthy population with unrestricted habitat, the chance that two carriers will mate is low, and even then, only 25 percent of their offspring will inherit the disease allele from both parents. While it is likely to happen at some point, it will not happen frequently enough for natural selection to be able to swiftly eliminate the allele from the population, and as a result, the allele will be maintained at low levels in the gene pool. However, if a family of carriers begins to interbreed with each other, this will dramatically increase the likelihood of two carriers mating and eventually producing diseased offspring, a phenomenon known as inbreeding depression.

Changes in allele frequencies that are identified in a population can shed light on how it is evolving. In addition to natural selection, there are other evolutionary forces that could be in play: natural selection, genetic drift, gene flow, mutation, and nonrandom mating.

Natural Selection

The theory of natural selection stems from the observation that some individuals in a population are more likely to survive longer and have more offspring than others; thus, they will pass on more of their genes to the next generation. A big, powerful male gorilla, for example, is much more likely than a smaller, weaker one to become the population's silverback, the pack's leader who mates far more than the other males of the group. The pack leader will father more offspring, who share half of his genes, and are likely to also grow bigger and stronger like their father. Over time, the genes for bigger size will increase in frequency in the population, and the population will, as a result, grow larger on average. That is, this would occur if this particular **selection pressure**, or driving selective force, were the only one acting on the population. In other examples, better camouflage or a stronger resistance to drought might pose a selection pressure.

Genetic Drift

Another way a population's allele and genotype frequencies can change is **genetic drift** (**Figure 16.3**), which is simply the effect of chance. By chance, some individuals will have more offspring than others—not due to an advantage conferred by some genetically-encoded trait, but just because one male happened to be in the right place at the right time (when the receptive female walked by) or because the other one happened to be in the wrong place at the wrong time (when a fox was hunting).

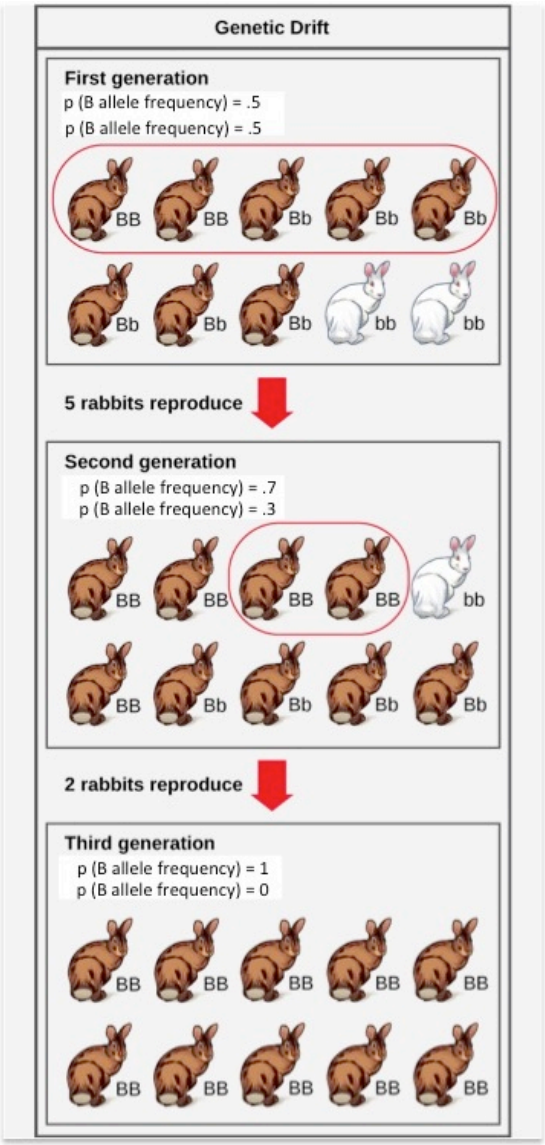


Figure 16.3 Genetic drift in a population can lead to the elimination of an allele from a population by chance. In this example, rabbits with the brown coat color allele (*B*) are dominant over rabbits with the white coat color allele (*b*). In the first generation, the two alleles occur with equal frequency in the population, resulting in p and q values of .5. Only half of the individuals reproduce, resulting in a second generation with p and q values of .7 and .3, respectively. Only two individuals in the second generation reproduce, and by chance these individuals are homozygous dominant for brown coat color. As a result, in the third generation the recessive *b* allele is lost.

Small populations are more susceptible to the forces of genetic drift. Large populations, on the other hand, are buffered against the effects of chance. If one individual of a population of 10 individuals happens to die at a young age before it leaves any offspring to the next generation, all of its genes—1/10 of the population's gene pool—will be suddenly lost. In a population of 100, that's only 1 percent of the overall gene pool; therefore, it is much less impactful on the population's genetic structure.

Genetic drift can also be magnified by natural events, such as a natural disaster that kills—at random—a large portion of the population. Known as the **bottleneck effect**, it results in a large portion of the genome suddenly being wiped out (**Figure 16.4**). In one fell swoop, the genetic structure of the survivors becomes the genetic structure of the entire population, which may be very different from the pre-disaster population.

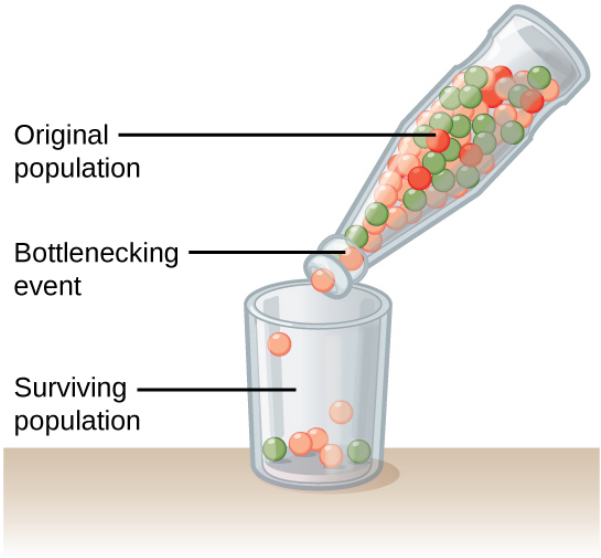


Figure 16.4 A chance event or catastrophe can reduce the genetic variability within a population.

Another scenario in which populations might experience a strong influence of genetic drift is if some portion of the population leaves to start a new population in a new location or if a population gets divided by a physical barrier of some kind. In this situation, those individuals are unlikely to be representative of the entire population, which results in the founder effect. The founder effect occurs when the genetic structure changes to match that of the new population's founding fathers and mothers. The founder effect is believed to have been a key factor in the genetic history of the Afrikaner population of Dutch settlers in South Africa, as evidenced by mutations that are common in Afrikaners but rare in most other populations. This is likely due to the fact that a higher-than-normal proportion of the founding colonists carried these mutations. As a result, the population expresses unusually high incidences of Huntington's disease (HD) and Fanconi anemia (FA), a genetic disorder known to cause blood marrow and congenital abnormalities—even cancer.[2]

Gene Flow

Another important evolutionary force is **gene flow**: the flow of alleles in and out of a population due to the migration of individuals or gametes (**Figure 16.5**). While some populations are fairly stable, others experience more flux. Many plants, for example, send their pollen far and wide, by wind or by bird, to pollinate other populations of the same species some distance away. Even a population that may initially appear to be stable, such as a pride of lions, can experience its fair share of immigration and emigration as developing males leave their mothers to seek out a new pride with genetically unrelated females. This variable flow of individuals in and out of the group not only changes the gene structure of the population, but it can also introduce new genetic variation to populations in different geological locations and habitats.

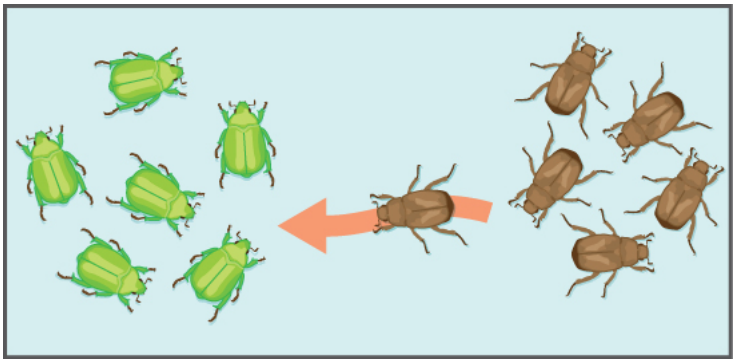


Figure 16.5 Gene flow can occur when an individual travels from one geographic location to another.

Mutation

Mutations are changes to an organism's DNA and are an important driver of diversity in populations. Species evolve

2. A. J. Tipping et al., "Molecular and Genealogical Evidence for a Founder Effect in Fanconi Anemia Families of the Afrikaner Population of South Africa," *PNAS* 98, no. 10 (2001): 5734-5739, doi: 10.1073/pnas.091402398.

because of the accumulation of mutations that occur over time. The appearance of new mutations is the most common way to introduce novel genotypic and phenotypic variance. Some mutations are unfavorable or harmful and are quickly eliminated from the population by natural selection. Others are beneficial and will spread through the population. Whether or not a mutation is beneficial or harmful is determined by whether it helps an organism survive to sexual maturity and reproduce. Some mutations do not do anything and can linger, unaffected by natural selection, in the genome. Some can have a dramatic effect on a gene and the resulting phenotype.

Nonrandom Mating

If individuals nonrandomly mate with their peers, the result can be a changing population. There are many reasons nonrandom mating occurs. One reason is mate choice; for example, female peahens may prefer peacocks with bigger, brighter tails. Traits that lead to more matings for an individual become selected for by natural selection. One common form of mate choice, called assortative mating, is an individual's preference to mate with partners who are phenotypically similar to themselves.

Another cause of nonrandom mating is physical location. This is especially true in large populations spread over large geographic distances where not all individuals will have equal access to one another. Some might be miles apart through woods or over rough terrain, while others might live immediately nearby.

16.3 | Formation of New Species

Introduction

> " As buds give rise by growth to fresh buds, and these, if vigorous, branch out and overtop on all sides many a feebler branch, so by generation I believe it has been with the great Tree of Life, which fills with its dead and broken branches the crust of the earth, and covers the surface with its ever branching and beautiful ramifications. "

Charles Darwin, *On the Origin of Species by Means of Natural Selection; or, The Preservation of Favoured Races in the Struggle for Life*, 1859

Darwin's insight about the branching tree of life is the easiest way to think about the origin of species. The twigs at the ends of branches may be different, but they all are connected and related. Although all life on earth shares various genetic similarities, only certain organisms combine genetic information by sexual reproduction and have offspring that can then successfully reproduce. Scientists call such organisms members of the same biological species.

Species and the Ability to Reproduce

A **species** is a group of individual organisms that interbreed and produce fertile, viable offspring. According to this definition, one species is distinguished from another when, in nature, it is not possible for matings between individuals from each species to produce fertile offspring. This idea of a species is called the **biological species concept** and is one of the most widely accepted species concepts.

Members of the same species share both external and internal characteristics, which develop from their DNA. The closer relationship two organisms share, the more DNA they have in common, just like people and their families. Your DNA is likely to be more like your father or mother's DNA than your cousin or grandparent's DNA. Organisms of the same species have the highest level of DNA alignment and therefore share characteristics and behaviors that lead to successful reproduction.

Species' appearance can be misleading in suggesting an ability or inability to mate. For example, even though domestic dogs (*Canis lupus familiaris*) display phenotypic differences, such as size, build, and coat, most dogs can interbreed and produce viable puppies that can mature and sexually reproduce (**Figure 16.6**).

(a) (b) (c)

Figure 16.6 The (a) poodle and (b) cocker spaniel can reproduce to produce a breed known as (c) the cockapoo. (credit a: modification of work by Sally Eller, Tom Reese; credit b: modification of work by Jeremy McWilliams; credit c: modification of work by Kathleen Conklin)

In other cases, individuals may appear similar although they are not members of the same species. For example, even though bald eagles (*Haliaeetus leucocephalus*) and African fish eagles (*Haliaeetus vocifer*) are both birds and eagles, each belongs to a separate species group (**Figure 16.7**). If humans were to artificially intervene and fertilize the egg of a bald eagle with the sperm of an African fish eagle and a chick did hatch, that offspring, called a **hybrid** (a cross between two species), would probably be infertile—unable to successfully reproduce after it reached maturity. Different species may have different genes that are active in development; therefore, it may not be possible to develop a viable offspring with two different sets of directions. Thus, even though hybridization may take place, the two species still remain separate.

(a) (b)

Figure 16.7 The (a) African fish eagle is similar in appearance to the (b) bald eagle, but the two birds are members of different species. (credit a: modification of work by Nigel Wedge; credit b: modification of work by U.S. Fish and Wildlife Service)

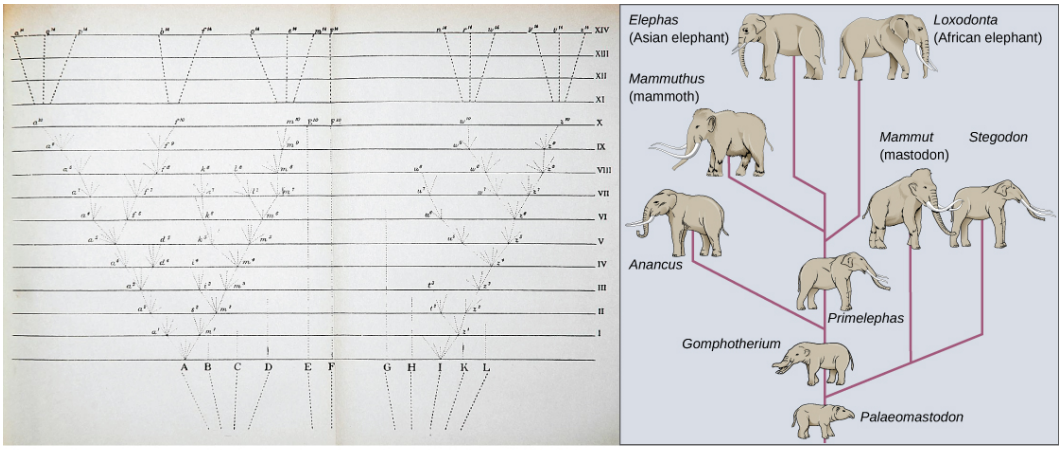
Populations of species share a gene pool: a collection of all the variants of genes in the species. Again, the basis to any changes in a group or population of organisms must be genetic for this is the only way to share and pass on traits. When variations occur within a species, they can only be passed to the next generation along two main pathways: asexual reproduction or sexual reproduction. The change will be passed on asexually simply if the reproducing cell possesses the changed trait. For the changed trait to be passed on by sexual reproduction, a gamete, such as a sperm or egg cell, must possess the changed trait. In other words, sexually-reproducing organisms can experience several genetic changes in their body cells, but if these changes do not occur in a sperm or egg cell, the changed trait will never reach the next generation. Only heritable traits can evolve. Therefore, reproduction plays a paramount role for genetic change to take root in a population or species. In short, organisms must be able to reproduce with each other to pass new traits to offspring.

Speciation

The biological definition of species, which works for sexually reproducing organisms, is a group of actually or potentially interbreeding individuals. There are exceptions to this rule. Many species are similar enough that hybrid offspring are possible and may often occur in nature, but for the majority of species this rule generally holds. In fact, the presence in nature of hybrids between similar species suggests that they may have descended from a single interbreeding species, and the speciation process may not yet be completed.

Given the extraordinary diversity of life on the planet there must be mechanisms for **speciation**: the formation of two species from one original species. Darwin envisioned this process as a branching event and diagrammed the process in the only illustration found in *On the Origin of Species* (**Figure 16.8a**). Compare this illustration to the diagram of elephant evolution (**Figure 16.8b**), which shows that as one species changes over time, it branches to form more than one new

species, repeatedly, as long as the population survives or until the organism becomes extinct.



(a) (b)

Figure 16.8 The only illustration in Darwin's *On the Origin of Species* is (a) a diagram showing speciation events leading to biological diversity. The diagram shows similarities to phylogenetic charts that are drawn today to illustrate the relationships of species. (b) Modern elephants evolved from the *Palaeomastodon*, a species that lived in Egypt 35–50 million years ago.

For speciation to occur, two new populations must be formed from one original population and they must evolve in such a way that it becomes impossible for individuals from the two new populations to interbreed. Biologists have proposed mechanisms by which this could occur that fall into two broad categories. **Allopatric speciation** (allo- = "other"; -patric = "homeland") involves geographic separation of populations from a parent species and subsequent evolution. **Sympatric speciation** (sym- = "same"; -patric = "homeland") involves speciation occurring within a parent species remaining in one location.

Biologists think of speciation events as the splitting of one ancestral species into two descendant species. There is no reason why there might not be more than two species formed at one time except that it is less likely and multiple events can be conceptualized as single splits occurring close in time.

Allopatric Speciation

A geographically continuous population has a gene pool that is relatively homogeneous. Gene flow, the movement of alleles across the range of the species, is relatively free because individuals can move and then mate with individuals in their new location. Thus, the frequency of an allele at one end of a distribution will be similar to the frequency of the allele at the other end. When populations become geographically discontinuous, that free-flow of alleles is prevented. When that separation lasts for a period of time, the two populations are able to evolve along different trajectories. Thus, their allele frequencies at numerous genetic loci gradually become more and more different as new alleles independently arise by mutation in each population. Typically, environmental conditions, such as climate, resources, predators, and competitors for the two populations will differ causing natural selection to favor divergent adaptations in each group.

Isolation of populations leading to allopatric speciation can occur in a variety of ways: a river forming a new branch, erosion forming a new valley, a group of organisms traveling to a new location without the ability to return, or seeds floating over the ocean to an island. The nature of the geographic separation necessary to isolate populations depends entirely on the biology of the organism and its potential for dispersal. If two flying insect populations took up residence in separate nearby valleys, chances are, individuals from each population would fly back and forth continuing gene flow. However, if two rodent populations became divided by the formation of a new lake, continued gene flow would be unlikely; therefore, speciation would be more likely.

Biologists group allopatric processes into two categories: dispersal and vicariance. Dispersal is when a few members of a species move to a new geographical area, and vicariance is when a natural situation arises to physically divide organisms.

Scientists have documented numerous cases of allopatric speciation taking place. For example, along the west coast of the United States, two separate sub-species of spotted owls exist. The northern spotted owl has genetic and phenotypic differences from its close relative: the Mexican spotted owl, which lives in the south (**Figure 16.9**).

Figure 16.9 The northern spotted owl and the Mexican spotted owl inhabit geographically separate locations with different climates and ecosystems. The owl is an example of allopatric speciation. (credit "northern spotted owl": modification of work by John and Karen Hollingsworth; credit "Mexican spotted owl": modification of work by Bill Radke)

Additionally, scientists have found that the further the distance between two groups that once were the same species, the more likely it is that speciation will occur. This seems logical because as the distance increases, the various environmental factors would likely have less in common than locations in close proximity. Consider the two owls: in the north, the climate is cooler than in the south; the types of organisms in each ecosystem differ, as do their behaviors and habits; also, the hunting habits and prey choices of the southern owls vary from the northern owls. These variances can lead to evolved differences in the owls, and speciation likely will occur.

Adaptive Radiation

In some cases, a population of one species disperses throughout an area, and each finds a distinct niche or isolated habitat. Over time, the varied demands of their new lifestyles lead to multiple speciation events originating from a single species. This is called **adaptive radiation** because many adaptations evolve from a single point of origin; thus, causing the species to radiate into several new ones. Island archipelagos like the Hawaiian Islands provide an ideal context for adaptive radiation events because water surrounds each island which leads to geographical isolation for many organisms. The Hawaiian honeycreeper illustrates one example of adaptive radiation. From a single species, called the founder species, numerous species have evolved, including the six shown in **Figure 16.10**.

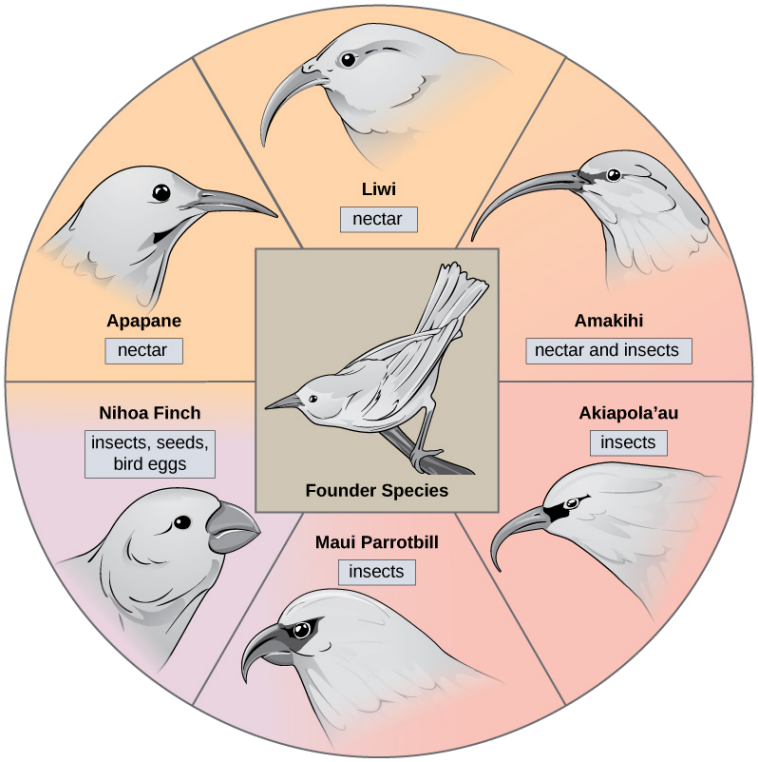


Figure 16.10 The honeycreeper birds illustrate adaptive radiation. From one original species of bird, multiple others evolved, each with its own distinctive characteristics.

Notice the differences in the species' beaks in **Figure 16.10**. Evolution in response to natural selection based on specific food sources in each new habitat led to evolution of a different beak suited to the specific food source. The seed-eating bird has a thicker, stronger beak which is suited to break hard nuts. The nectar-eating birds have long beaks to dip into flowers to reach the nectar. The insect-eating birds have beaks like swords, appropriate for stabbing and impaling insects. Darwin's finches are another example of adaptive radiation in an archipelago.

Sympatric Speciation

Can divergence occur if no physical barriers are in place to separate individuals who continue to live and reproduce in the same habitat? The answer is yes. The process of speciation within the same space is called sympatric speciation; the prefix "sym" means same, so "sympatric" means "same homeland" in contrast to "allopatric" meaning "other homeland." A number of mechanisms for sympatric speciation have been proposed and studied.

One form of sympatric speciation can begin with a serious chromosomal error during cell division. In a normal cell division event chromosomes replicate, pair up, and then separate so that each new cell has the same number of chromosomes. However, sometimes the pairs separate and the end cell product has too many or too few individual chromosomes in a condition called aneuploidy (**Figure 16.11**).

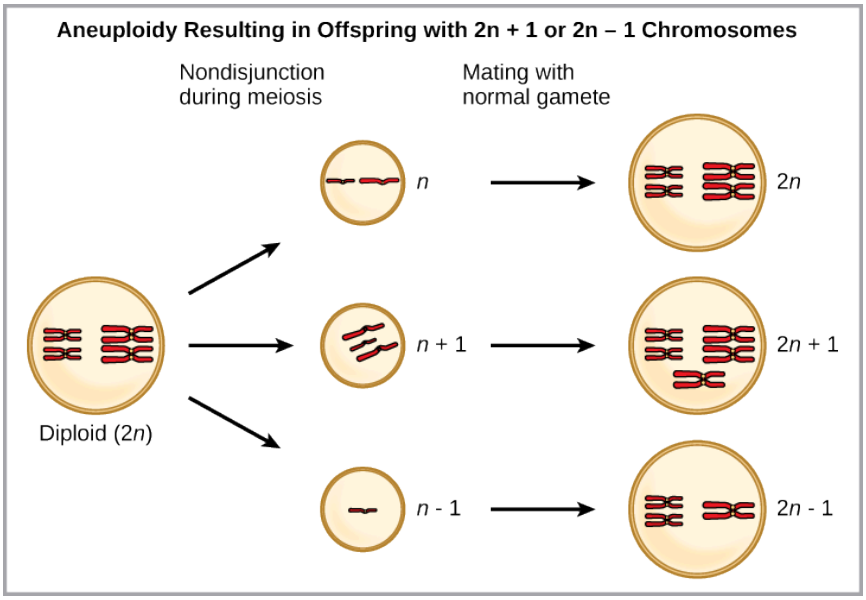


Figure 16.11 Aneuploidy results when the gametes have too many or too few chromosomes due to nondisjunction during meiosis. In the example shown here, the resulting offspring will have 2*n*+1 or 2*n*-1 chromosomes

Polyploidy is a condition in which a cell or organism has an extra set, or sets, of chromosomes. Scientists have identified two main types of polyploidy that can lead to reproductive isolation of an individual in the polyploidy state. Reproductive isolation is the inability to interbreed. In some cases, a polyploid individual will have two or more complete sets of chromosomes from its own species in a condition called autopolyploidy (**Figure 16.12**). The prefix "auto-" means "self," so the term means multiple chromosomes from one's own species. Polyploidy results from an error in meiosis in which all of the chromosomes move into one cell instead of separating.

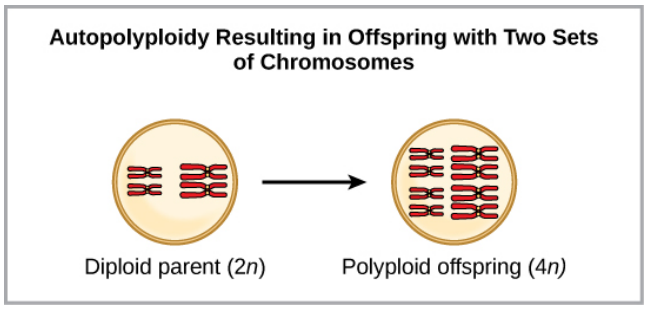


Figure 16.12 Autopolyploidy results when mitosis is not followed by cytokinesis.

For example, if a plant species with $2n = 6$ produces autopolyploid gametes that are also diploid ($2n = 6$, when they should be $n = 3$), the gametes now have twice as many chromosomes as they should have. These new gametes will be incompatible with the normal gametes produced by this plant species. However, they could either self-pollinate or reproduce with other autopolyploid plants with gametes having the same diploid number. In this way, sympatric speciation can occur quickly by forming offspring with 4*n* called a tetraploid. These individuals would immediately be able to reproduce only with those of this new kind and not those of the ancestral species.

The other form of polyploidy occurs when individuals of two different species reproduce to form a viable offspring called an allopolyploid. The prefix "allo-" means "other" (recall from allopatric): therefore, an allopolyploid occurs when gametes from two different species combine. **Figure 16.13** illustrates one possible way an allopolyploid can form. Notice how it takes two generations, or two reproductive acts, before the viable fertile hybrid results.

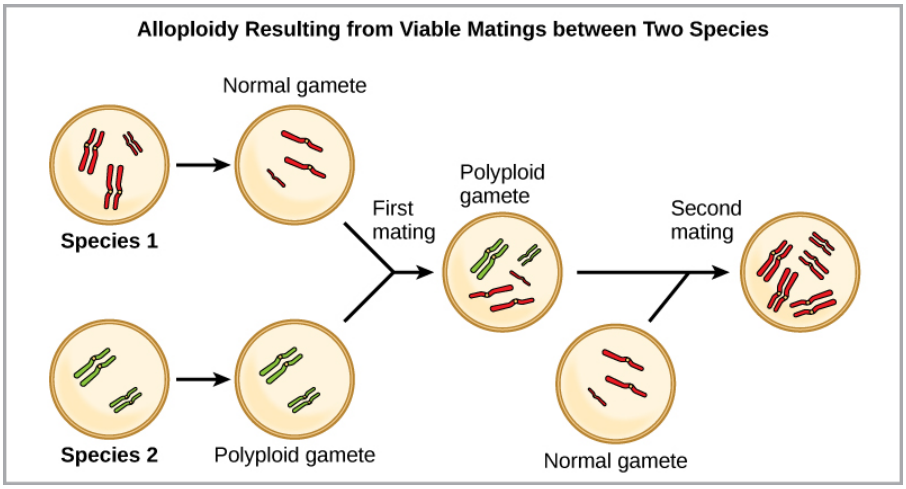


Figure 16.13 Alloploidy results when two species mate to produce viable offspring. In the example shown, a normal gamete from one species fuses with a polyploidy gamete from another. Two matings are necessary to produce viable offspring.

The cultivated forms of wheat, cotton, and tobacco plants are all allopolyploids. Although polyploidy occurs occasionally in animals, it takes place most commonly in plants. (Animals with any of the types of chromosomal aberrations described here are unlikely to survive and produce normal offspring.) Scientists have discovered more than half of all plant species studied relate back to a species evolved through polyploidy. With such a high rate of polyploidy in plants, some scientists hypothesize that this mechanism takes place more as an adaptation than as an error.

Reproductive Isolation

Given enough time, the genetic and phenotypic divergence between populations will affect characters that influence reproduction: if individuals of the two populations were to be brought together, mating would be less likely, but if mating occurred, offspring would be non-viable or infertile. Many types of diverging characters may affect the **reproductive isolation**, the ability to interbreed, of the two populations.

Reproductive isolation can take place in a variety of ways. Scientists organize them into two groups: prezygotic barriers and postzygotic barriers. Recall that a zygote is a fertilized egg: the first cell of the development of an organism that reproduces sexually. Therefore, a **prezygotic barrier** is a mechanism that blocks reproduction or prevents fertilization when organisms attempt reproduction. A **postzygotic barrier** occurs after fertilization; this includes organisms that don't survive the embryonic stage and those that are born sterile.

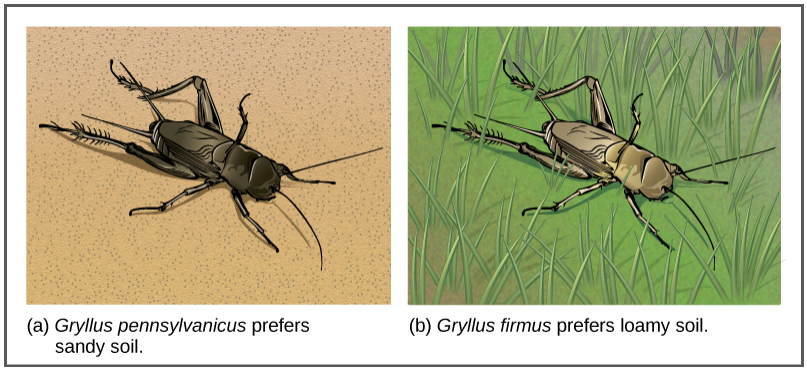
Some types of prezygotic barriers prevent reproduction entirely. Many organisms only reproduce at certain times of the year, often just annually. Differences in breeding schedules, called **temporal isolation**, can act as a form of reproductive isolation. For example, two species of frogs inhabit the same area, but one reproduces from January to March, whereas the other reproduces from March to May (**Figure 16.14**).

(a) (b)

Figure 16.14 These two related frog species exhibit temporal reproductive isolation. (a) *Rana aurora* breeds earlier in the year than (b) *Rana boylii*. (credit a: modification of work by Mark R. Jennings, USFWS; credit b: modification of work by Alessandro Catenazzi)

In some cases, populations of a species move or are moved to a new habitat and take up residence in a place that no longer

overlaps with the other populations of the same species. This situation is called **habitat isolation**. Reproduction with the parent species ceases, and a new group exists that is now reproductively and genetically independent. For example, a cricket population that was divided after a flood could no longer interact with each other. Over time, the forces of natural selection, mutation, and genetic drift will likely result in the divergence of the two groups (**Figure 16.15**).



(a) *Gryllus pennsylvanicus* prefers sandy soil.

(b) *Gryllus firmus* prefers loamy soil.

Figure 16.15 Speciation can occur when two populations occupy different habitats. The habitats need not be far apart. The cricket (a) *Gryllus pennsylvanicus* prefers sandy soil, and the cricket (b) *Gryllus firmus* prefers loamy soil. The two species can live in close proximity, but because of their different soil preferences, they became genetically isolated.

Behavioral isolation occurs when the presence or absence of a specific behavior prevents reproduction from taking place. For example, male fireflies use specific light patterns to attract females. Various species of fireflies display their lights differently. If a male of one species tried to attract the female of another, she would not recognize the light pattern and would not mate with the male.

Other prezygotic barriers work when differences in their gamete cells (eggs and sperm) prevent fertilization from taking place; this is called a **gametic barrier**. Similarly, in some cases closely related organisms try to mate, but their reproductive structures simply do not fit together. For example, damselfly males of different species have differently shaped reproductive organs. If one species tries to mate with the female of another, their body parts simply do not fit together. (**Figure 16.16**).

Figure 16.16 The shape of the male reproductive organ varies among male damselfly species, and is only compatible with the female of that species. Reproductive organ incompatibility keeps the species reproductively isolated.

In plants, certain structures aimed to attract one type of pollinator simultaneously prevent a different pollinator from accessing the pollen. The tunnel through which an animal must access nectar can vary widely in length and diameter, which prevents the plant from being cross-pollinated with a different species (**Figure 16.17**).

(a) Honeybee drinking nectar from a foxglove flower

(b) Ruby-throated hummingbird drinking nectar from a trumpet creeper flower

Figure 16.17 Some flowers have evolved to attract certain pollinators. The (a) wide foxglove flower is adapted for pollination by bees, while the (b) long, tube-shaped trumpet creeper flower is adapted for pollination by humming birds.

When fertilization takes place and a zygote forms, postzygotic barriers can prevent reproduction. Hybrid individuals in many cases cannot form normally in the womb and simply do not survive past the embryonic stages. This is called **hybrid inviability** because the hybrid organisms simply are not viable. In another postzygotic situation, reproduction leads to the

birth and growth of a hybrid that is sterile and unable to reproduce offspring of their own; this is called hybrid sterility.

Habitat Influence on Speciation

Sympatric speciation may also take place in ways other than polyploidy. For example, consider a species of fish that lives in a lake. As the population grows, competition for food also grows. Under pressure to find food, suppose that a group of these fish had the genetic flexibility to discover and feed off another resource that was unused by the other fish. What if this new food source was found at a different depth of the lake? Over time, those feeding on the second food source would interact more with each other than the other fish; therefore, they would breed together as well. Offspring of these fish would likely behave as their parents: feeding and living in the same area and keeping separate from the original population. If this group of fish continued to remain separate from the first population, eventually sympatric speciation might occur as more genetic differences accumulated between them.

This scenario does play out in nature, as do others that lead to reproductive isolation. One such place is Lake Victoria in Africa, famous for its sympatric speciation of cichlid fish. Researchers have found hundreds of sympatric speciation events in these fish, which have not only happened in great number, but also over a short period of time. **Figure 16.18** shows this type of speciation among a cichlid fish population in Nicaragua. In this locale, two types of cichlids live in the same geographic location but have come to have different morphologies that allow them to eat various food sources.

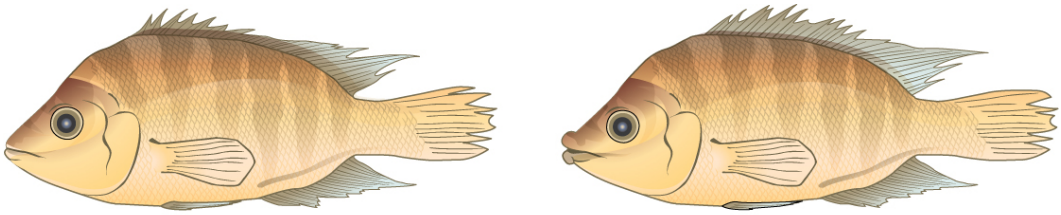


Figure 16.18 Cichlid fish from Lake Apoyeque, Nicaragua, show evidence of sympatric speciation. Lake Apoyeque, a crater lake, is 1800 years old, but genetic evidence indicates that the lake was populated only 100 years ago by a single population of cichlid fish. Nevertheless, two populations with distinct morphologies and diets now exist in the lake, and scientists believe these populations may be in an early stage of speciation.

17 | CELLULAR ENERGETICS

17.1 | Energy and Metabolism

Introduction

> “ [In research on bacteria metabolism] we have indeed much the same position as an observer trying to gain an idea of the life of a household by careful scrutiny of the persons and material arriving or leaving the house; we keep accurate records of the foods and commodities left at the door and patiently examine the contents of the dust-bin and endeavour to deduce from such data the events occurring within the closed doors. ”

Marjory Stephenson (1930)

Metabolism is indeed most closely studied by marking what goes in and what comes out, both in terms of materials and energy. Scientists use the term **bioenergetics** to discuss the concept of energy flow (**Figure 17.1**) through living systems from ecosystems to 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 what has been used, cells must continually produce more energy to replenish that used by the many energy-requiring chemical reactions that constantly take place. All of the chemical reactions that take place inside cells, including those that use energy and those that release energy, are the cell’s **metabolism**.

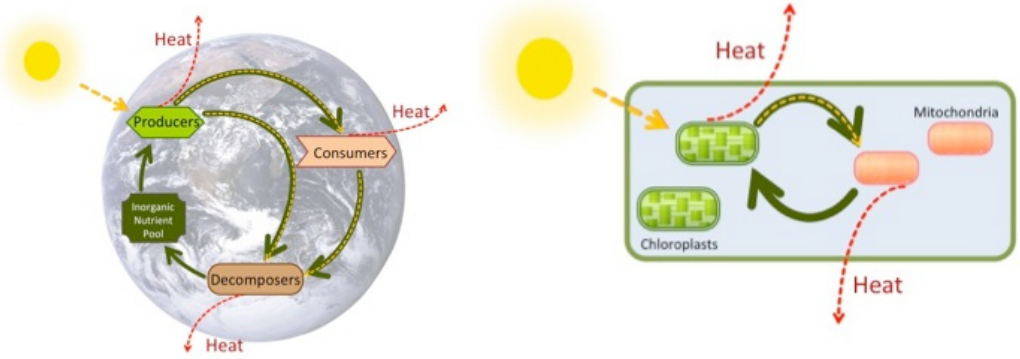


Figure 17.1 Most life forms on earth get their energy from the sun. Various metabolic processes harvest and harness this energy which ultimately leaves the system as heat (2nd Law of Thermodynamics). The dashed lines represent energy and the solid lines represent nutrients. (Image by Eva Horne and Robert Bear)

Metabolism of Carbohydrates

The metabolism of sugar (a simple carbohydrate) is a classic example of the many cellular processes that use and produce energy. Living things consume sugar as a major energy source, because sugar molecules have a great deal of energy stored

within their bonds. The breakdown of glucose, a simple sugar, is described by the equation:

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

Carbohydrates that are consumed have their origins in photosynthesizing organisms like plants (**Figure 17.2**). During photosynthesis, plants use the energy of sunlight to convert carbon dioxide gas (CO_2) into sugar molecules, like glucose ($C_6H_{12}O_6$). Because this process involves synthesizing a larger, energy-storing molecule, it requires an input of energy to proceed. The synthesis of glucose is described by this equation (notice that it is the reverse of the previous equation):

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

During the chemical reactions of photosynthesis, energy is provided in the form of a very high-energy molecule called ATP, or adenosine triphosphate, 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. The sugar (glucose) is stored as starch or glycogen. Energy-storing polymers like these are broken down into glucose to supply molecules of ATP.

Solar energy is required to synthesize a molecule of glucose during the reactions of photosynthesis. In photosynthesis, light energy from the sun is initially transformed into chemical energy that is temporally stored in the energy carrier molecules ATP and NADPH (nicotinamide adenine dinucleotide phosphate). The stored energy in ATP and NADPH is then used later in photosynthesis to build one molecule of glucose from six molecules of CO_2. This process is analogous to eating breakfast in the morning to acquire energy for your body that can be used later in the day. Under ideal conditions, energy from 18 molecules of ATP is required to synthesize one molecule of glucose during the reactions of photosynthesis. Glucose molecules can also be combined with and converted into other types of sugars. When sugars are consumed, molecules of glucose eventually make their way into each living cell of the organism. Inside the cell, each sugar molecule is broken down through a complex series of chemical reactions. The goal of these reactions is to harvest the energy stored inside the sugar molecules. The harvested energy is used to make high-energy ATP molecules, which can be used to perform work, powering many chemical reactions in the cell. The amount of energy needed to make one molecule of glucose from six molecules of carbon dioxide is 18 molecules of ATP and 12 molecules of NADPH (each one of which is energetically equivalent to three molecules of ATP), or a total of 54 molecule equivalents required for the synthesis of one molecule of glucose. This process is a fundamental and efficient way for cells to generate the molecular energy that they require.

Figure 17.2 Plants, like this oak tree and acorn, use energy from sunlight to make sugar and other organic molecules. Both plants and animals (like this squirrel) use cellular respiration to derive energy from the organic molecules originally produced by plants. (credit “acorn”: modification of work by Noel Reynolds; credit “squirrel”: modification of work by Dawn Huczek)

Metabolic Pathways

The processes of making and breaking down sugar molecules illustrate two types of metabolic pathways. A metabolic pathway is a series of interconnected biochemical reactions that convert a substrate molecule or molecules, step-by-step, through a series of metabolic intermediates, eventually yielding a final product or products. In the case 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** (building) and **catabolic** (breaking down) pathways, respectively. Consequently, metabolism is composed of building (anabolism) and degradation (catabolism).

evolution CONNECTION

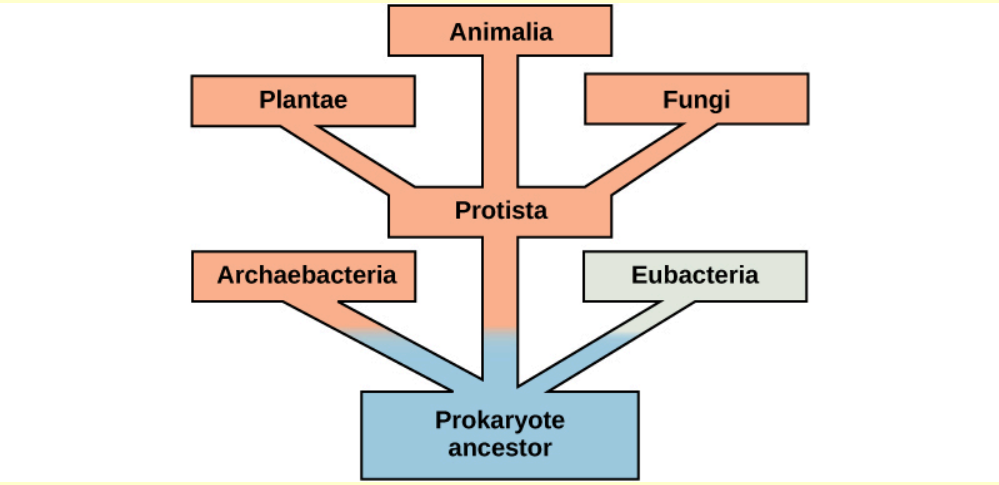


Figure 17.3 This tree shows the evolution of the various branches of life. The vertical dimension is time. Early life forms, in blue, used anaerobic metabolism to obtain energy from their surroundings.

Evolution of Metabolic Pathways

There is more to the complexity of metabolism than understanding the metabolic pathways alone. Metabolic complexity varies from organism to organism. Photosynthesis is the primary pathway in which photosynthetic organisms like plants (the majority of global synthesis is done by planktonic algae) harvest the sun's energy and convert it into carbohydrates. The by-product of photosynthesis is oxygen, required by some cells to carry out cellular respiration. During cellular respiration, oxygen aids in the catabolic breakdown of carbon compounds, like carbohydrates. Among the products of this catabolism are CO_2 and ATP. In addition, some eukaryotes perform catabolic processes without oxygen (fermentation); that is, they perform or use anaerobic metabolism.

Organisms probably evolved anaerobic metabolism to survive (living organisms came into existence about 3.8 billion years ago, when the atmosphere lacked oxygen). Despite the differences between organisms and the complexity of metabolism, researchers have found that all branches of life share some of the same metabolic pathways, suggesting that all organisms evolved from the same ancient common ancestor (**Figure 17.3**). Evidence indicates that over time, the pathways diverged, adding specialized enzymes to allow organisms to better adapt to their environment, thus increasing their chance to survive. However, the underlying principle remains that all organisms must harvest energy from their environment and convert it to ATP to carry out cellular functions.

Anabolic and Catabolic Pathways

Anabolic pathways require an input of energy to synthesize complex molecules from simpler ones. Synthesizing sugar from CO_2 is one example. Other examples are the synthesis of large proteins from amino acid building blocks, and the synthesis of new DNA strands from nucleic acid building blocks. These biosynthetic processes are critical to the life of the cell, take place constantly, and demand energy provided by ATP and other high-energy molecules like NADH (nicotinamide adenine dinucleotide) and NADPH (**Figure 17.4**).

ATP is an important molecule for cells to have in sufficient supply at all times. The breakdown of sugars illustrates how a single molecule of glucose can store enough energy to make a great deal of ATP, 36 to 38 molecules. This is a **catabolic** pathway. Catabolic pathways involve the degradation (or breakdown) of complex molecules into simpler ones. Molecular energy stored in the bonds of complex molecules is released in catabolic pathways and harvested in such a way that it can be used to produce ATP. Other energy-storing molecules, such as fats, are also broken down through similar catabolic reactions to release energy and make ATP (**Figure 17.4**).

It is important to know that the chemical reactions of metabolic pathways don't take place spontaneously. 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.

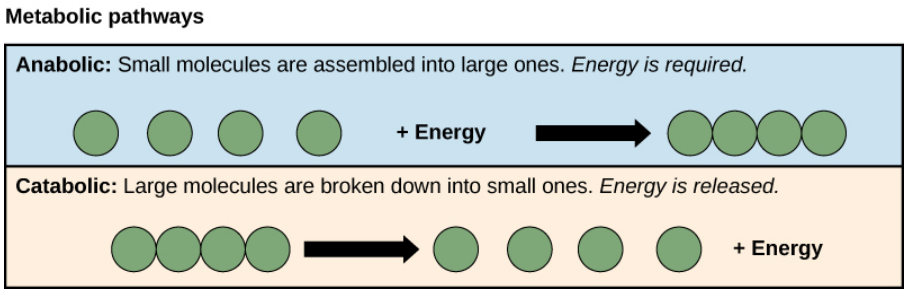


Figure 17.4 Anabolic pathways are those that require energy to synthesize larger molecules. Catabolic pathways are those that generate energy by breaking down larger molecules. Both types of pathways are required for maintaining the cell's energy balance.

17.2 | Thermodynamics

Introduction

> " A theory is the more impressive the greater the simplicity of its premises is, the more different kinds of things it relates, and the more extended is its area of applicability. Therefore the deep impression which classical thermodynamics made upon me. It is the only physical theory of universal content concerning which I am convinced that within the framework of the applicability of its basic concepts, it will never be overthrown. "

Albert Einstein (1946)

That high praise from Einstein still rings true today; the laws of thermodynamics can still explain many interactions, both in the biotic and abiotic realms. **Thermodynamics** refers to the study of energy and energy transfer involving physical matter. The matter and its environment relevant to a particular case of energy transfer are classified as a system, and everything outside of that system 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. An open system is one in which energy can be transferred between the system and its surroundings. The stovetop system is open because heat can be lost into the air. A closed system is one that cannot transfer energy to its surroundings.

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

The First Law of Thermodynamics

The first law of thermodynamics deals with the total amount of energy in the universe. It states that this total amount of energy in the universe is constant. 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 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 into the chemical energy stored within organic molecules. Some examples of energy transformations are shown in **Figure 17.5**.

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 very well. Chemical energy stored within organic molecules such as sugars and fats is 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 beating motion of cilia or flagella, contracting muscle fibers to create movement, and reproduction.

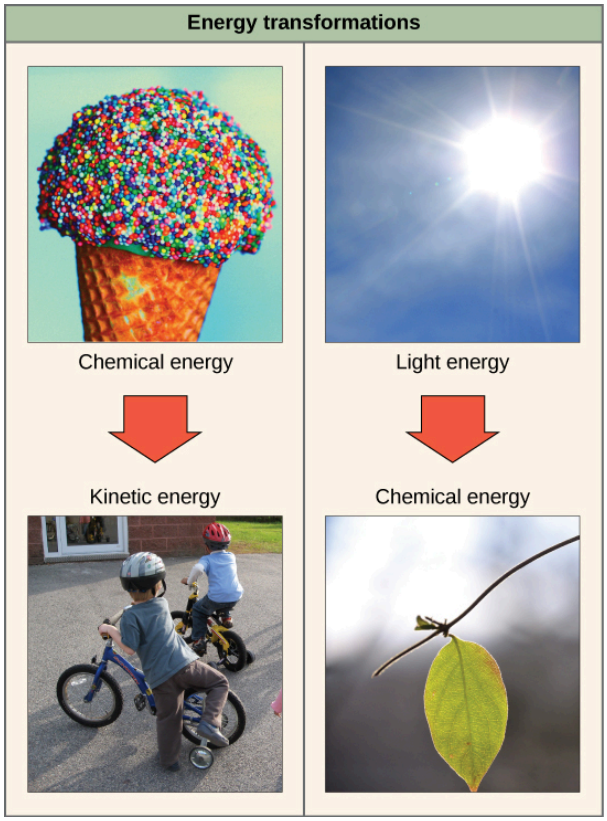


Figure 17.5 Shown are two examples of energy being transferred from one system to another and transformed from one form to another. Humans can convert the chemical energy in food, like this ice cream cone, into kinetic energy (the energy of movement to ride a bicycle). Plants can convert electromagnetic radiation (light energy) from the sun into chemical energy. (credit “ice cream”: modification of work by D. Sharon Pruitt; credit “kids on bikes”: modification of work by Michelle Riggen-Ransom; credit “leaf”: modification of work by Cory Zanker)

The Second Law of Thermodynamics

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. None of the energy transfers we’ve discussed, along with all energy transfers and transformations in the universe, is 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 doing work. For example, when an airplane flies through the air, some of the energy of the flying plane is lost as heat energy due to friction with the surrounding air. This friction actually heats the air by temporarily increasing the speed of air molecules. Likewise, some energy is lost as heat energy during cellular metabolic reactions. This is good for warm-blooded creatures like us, because heat energy helps to maintain our body temperature. Strictly speaking, no energy transfer is completely efficient, because some energy is lost in an unusable form.

An important concept in physical systems is that of order and disorder (also known as randomness). 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 (**Figure 17.6**). To better understand entropy, think of a student’s bedroom. If no energy or work were put into it, the room would quickly become messy. It would exist in a very disordered state, one of high entropy. Energy must be put into the system, in the form of the student doing work and putting everything away, in order to bring the room back to a state of cleanliness and order. This state is one of low entropy. Similarly, a car or house must be constantly maintained with work in order to keep it in an ordered state. Left alone, the entropy of the house or car gradually increases through rust and degradation. Molecules and chemical reactions have varying amounts of entropy as well. For example, as chemical reactions reach a state of equilibrium,

entropy increases, and as molecules at a high concentration in one place diffuse and spread out, entropy also increases.

scientific method CONNECTION

Transfer of Energy and the Resulting Entropy

Set up a simple experiment to understand how energy is transferred and how a change in entropy results.

1. Take a block of ice. This is water in solid form, so it has a high structural order. This means that the molecules cannot move very much and are in a fixed position. The temperature of the ice is 0°C. As a result, the entropy of the system is low.
2. Allow the ice to melt at room temperature. What is the state of molecules in the liquid water now? How did the energy transfer take place? Is the entropy of the system higher or lower? Why?
3. Heat the water to its boiling point. What happens to the entropy of the system when the water is heated?

All physical systems can be thought of in this way: Living things are highly ordered, requiring constant energy input to be maintained in a state of low entropy. As living systems take in energy-storing molecules and transform them through chemical reactions, they lose some amount of usable energy in the process, because no reaction is completely efficient. They also produce waste and by-products that aren't useful energy sources. This process increases the entropy of the system's surroundings. Since all energy transfers result in the loss of some usable energy, the second law of thermodynamics states that every energy transfer or transformation increases the entropy of the universe. Even though living things are highly ordered and maintain a state of low entropy, the entropy of the universe in total is constantly increasing due to the loss of usable energy with each energy transfer that occurs. Essentially, living things are in a continuous uphill battle against this constant increase in universal entropy.

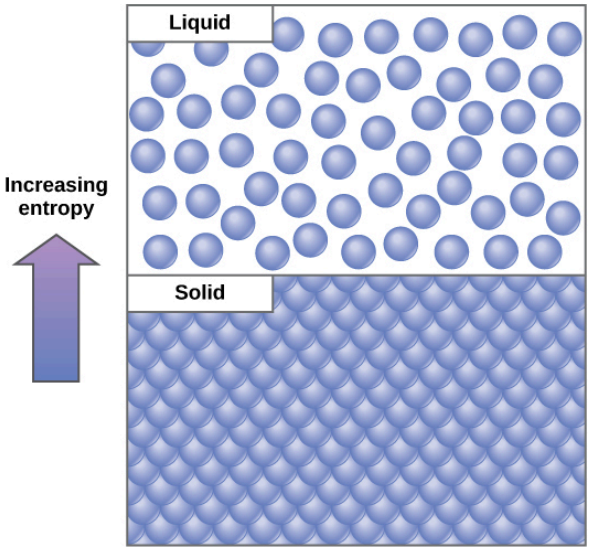


Figure 17.6 Entropy is a measure of randomness or disorder in a system. Gases have higher entropy than liquids, and liquids have higher entropy than solids.

17.3 | Potential, Kinetic, and Free Energy

Introduction

"Energie is the operation, efflux or activity of any being: as the light of the Sunne is the energie of the Sunne, and every phantasm of

the soul is the energie of the soul. ”

Henry More (1642), this is the first recorded definition of the term energy in English.

Energy is defined as the ability to do work. As you know, energy exists in different forms. For example, chemical energy, electrical energy, light energy, and heat energy are all different types of energy. While these are all familiar types of energy that one can see or feel, there is another type of energy that is much less tangible. This energy is associated with something as simple as an object held above the ground. In order to appreciate the way energy flows into and out of biological systems, it is important to understand more about the different types of energy that exist in the physical world.

Types of Energy

When an object is in motion, there is energy associated with that object. In the example of an airplane in flight, there is a great deal of energy associated with the motion of the airplane. This is because moving objects are capable of enacting a change, or doing work. Think of a wrecking ball. Even a slow-moving wrecking ball can do a great deal of damage to other objects. However, a wrecking ball that is not in motion is incapable of performing work. Energy associated with objects in motion is called **kinetic energy**. A speeding bullet, a walking person, the rapid movement of molecules in the air (which produces heat), and electromagnetic radiation like light all have kinetic energy.

Now what if that same motionless wrecking ball is lifted two stories above a car with a crane? If the suspended wrecking ball is unmoving, is there energy associated with it? The answer is yes. The suspended wrecking ball has energy associated with it that is fundamentally different from the kinetic energy of objects in motion. This form of energy results from the fact that there is the *potential* for the wrecking ball to do work. If it is released, indeed it would do work. Because this type of energy refers to the potential to do work, it is called **potential energy**. Objects transfer their energy between kinetic and potential in the following way: As the wrecking ball hangs motionless, it has 0 kinetic and 100 percent potential energy. Once it is released, its kinetic energy begins to increase because it builds speed due to gravity. At the same time, as it nears the ground, it loses potential energy. Somewhere mid-fall it has 50 percent kinetic and 50 percent potential energy. Just before it hits the ground, the ball has nearly lost its potential energy and has near-maximal kinetic energy. Other examples of potential energy include the energy of water held behind a dam (**Figure 17.7**), or a person about to skydive out of an airplane.

Figure 17.7 Water behind a dam 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 (such as a child sitting on a tree branch), but also with the structure of matter. A spring on the ground has potential energy if it is compressed; so does a rubber band that is pulled taut. The very existence of living cells relies heavily on structural potential energy. On a chemical level, the bonds that hold the atoms of molecules together have 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** (**Figure 17.8**). Chemical energy is responsible for providing living cells with energy from food. The release of energy is brought about by breaking

the molecular bonds within fuel molecules.

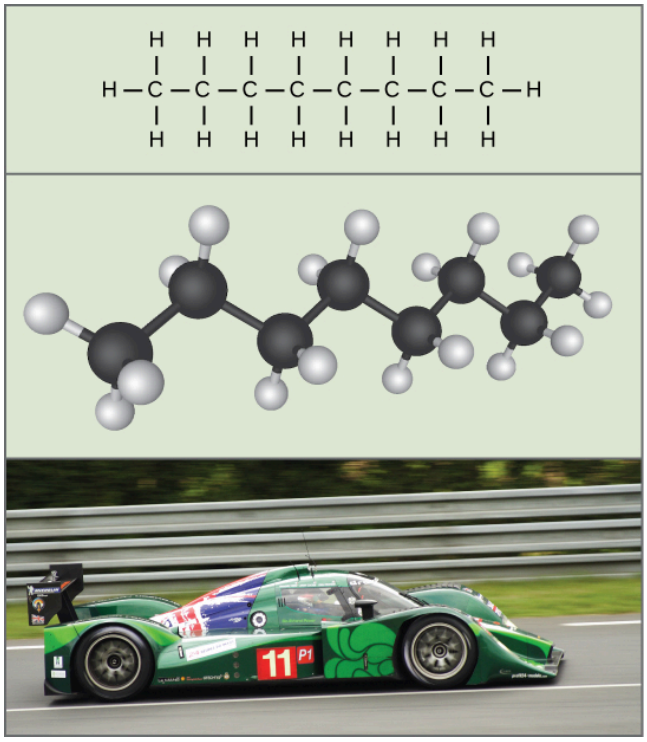


Figure 17.8 The molecules in gasoline (octane, the chemical formula shown) contain chemical energy within the chemical bonds. This energy is transformed into kinetic energy that allows a car to race on a racetrack. (credit "car": modification of work by Russell Trow)

Free Energy

After learning that chemical reactions release energy when energy-storing bonds are broken, an important next question is how is the energy associated with 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 quantitate these energy transfers. Free energy is called Gibbs free energy (abbreviated with the letter G) after Josiah Willard Gibbs, the scientist who developed the measurement. 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, resulting in entropy. Gibbs free energy specifically refers to the energy associated with a chemical reaction that is available after entropy is accounted for. In other words, Gibbs free energy is usable energy, or energy that is available to do work. So, every chemical reaction involves a change in free energy, called delta G (ΔG).

Endergonic Reactions and Exergonic Reactions

If energy is released during a chemical reaction, then the resulting value from the above equation will be a negative number. In other words, reactions that release energy have a $\Delta G < 0$. A negative ΔG also means that the products of the reaction have less free energy than the reactants, because they gave off some free energy during the reaction. Reactions that have a negative ΔG and consequently release free energy are called **exergonic reactions**. Think: *ex*ergonic means energy is *ex*iting the system. These reactions are also referred to as spontaneous reactions, because they can occur without the addition of energy into the system. Understanding which chemical reactions are spontaneous and release free energy is extremely useful for biologists, because these reactions can be harnessed to perform work inside the cell. An important distinction must be drawn between the term spontaneous and the idea of a chemical reaction that occurs 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 requires an input of energy rather than releasing energy, 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.

Let's revisit the example of the synthesis and breakdown of the food molecule, glucose. Remember that the building of complex molecules, such as sugars, from simpler ones is an anabolic process and requires energy. Therefore, the chemical

reactions involved in anabolic processes are endergonic reactions. On the other hand, the catabolic process of breaking sugar down into simpler molecules releases energy in a series of exergonic reactions. Like the example of rust above, the breakdown of sugar involves spontaneous reactions, but these reactions don't occur instantaneously. **Figure 17.9** shows some other examples of endergonic and exergonic reactions. Later sections will provide more information about what else is required to make even spontaneous reactions happen more efficiently.

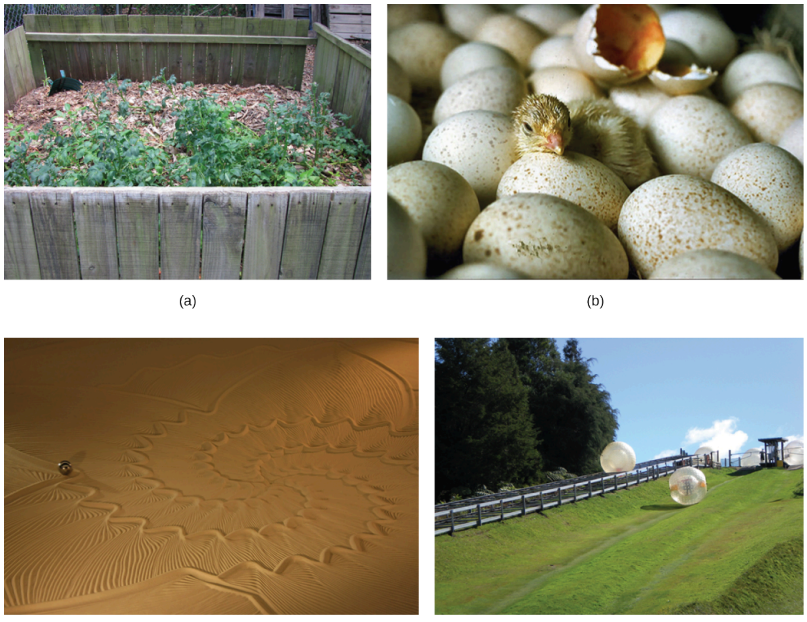


Figure 17.9 Shown are some examples of endergonic processes (ones that require energy) and exergonic processes (ones that release energy). These include (a) a compost pile decomposing, (b) a chick hatching from a fertilized egg, (c) sand art being destroyed, and (d) a ball rolling down a hill. (credit a: modification of work by Natalie Maynor; credit b: modification of work by USDA; credit c: modification of work by "Athlex"/Flickr; credit d: modification of work by Harry Malsch)

An important concept in the study of metabolism and energy is that of chemical equilibrium. Most chemical reactions are reversible. They can proceed in both directions, releasing energy into their environment in one direction, and absorbing it from the environment in the other direction (**Figure 17.10**). The same is true for the chemical reactions involved in cell metabolism, such as the breaking down and building up of proteins into and from individual amino acids, respectively. Reactants within a closed system will undergo chemical reactions in both directions until a state of equilibrium is reached. This state of equilibrium is one of the lowest possible free energy and a state of maximal entropy. Energy must be put into the system to push the reactants and products away from a state of equilibrium. Either reactants or products must be added, removed, or changed. If a cell were a closed system, its chemical reactions would reach equilibrium, and it would die because there would be insufficient free energy left to perform the work needed to maintain life. In a living cell, chemical reactions are constantly moving towards equilibrium, but never reach it. This is because a living cell is an open system. Materials pass in and out, the cell recycles the products of certain chemical reactions into other reactions, and chemical equilibrium is never reached. In this way, living organisms are in a constant energy-requiring, uphill battle against equilibrium and entropy. This constant supply of energy ultimately comes from sunlight, which is used to produce nutrients in the process of photosynthesis.

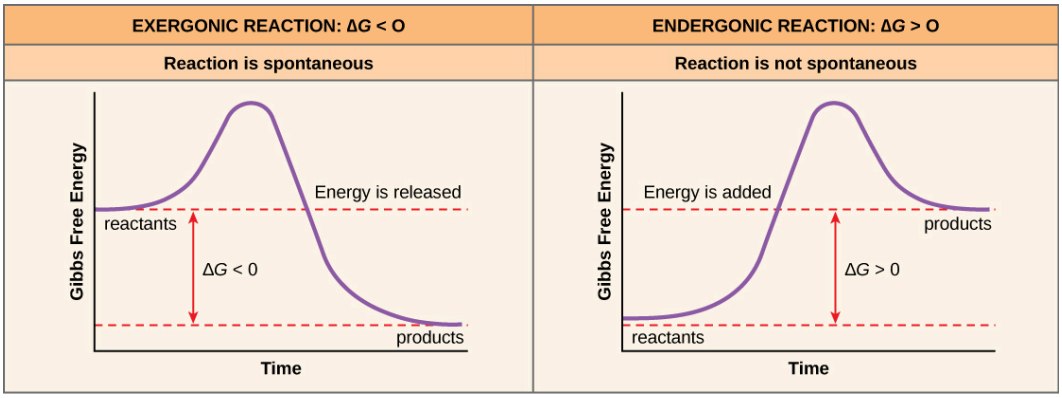


Figure 17.10 Exergonic and endergonic reactions result in changes in Gibbs free energy. Exergonic reactions release energy; endergonic reactions require energy to proceed.

17.4 | Energy in Living Systems

Introduction

> "There is deposited in them (plants) an enormous quantity of potential energy, whose equivalent is provided to us as heat in the burning of a plant substance. So far as we know at present, the only living energy absorbed during plant growth are the rays of sunlight... Animals take up oxygen and complex oxidizable compounds made by plants, release largely as combustion products carbonic acid and water, ...thus using a certain amount of chemical potential energy to produce heat and mechanical forces."

Hermann Ludwig Ferdinand von Helmholtz, 1847.

Energy production within a cell involves many coordinated chemical pathways. Most of these pathways are combinations of oxidation and reduction reactions. Oxidation and reduction occur in tandem. An **oxidation** reaction strips an electron from an atom in a compound, and the addition of this electron to another compound is a **reduction** reaction. Because oxidation and reduction usually occur together, these pairs of reactions are called oxidation reduction reactions, or **redox reactions**.

Electrons and Energy

The removal of an electron from a molecule, oxidizing it, results in a decrease in potential energy in the oxidized compound. The electron (sometimes as part of a hydrogen atom), does not remain unbonded, however, in the cytoplasm of a cell. Rather, the electron is shifted to a second compound, reducing the second compound. The shift of an electron from one compound to another removes some potential energy from the first compound (the oxidized compound) and increases the potential energy of the second compound (the reduced compound)(**Figure 17.11**). The transfer of electrons between molecules is important because most of the energy stored in atoms and used to fuel cell functions is in the form of high-energy electrons. The transfer of energy in the form of electrons allows the cell to transfer and use energy in an incremental fashion—in small packages rather than in a single, destructive burst. This section focuses on the extraction of energy from food; you will see that as you track the path of the transfers, you are tracking the path of electrons moving through metabolic pathways.

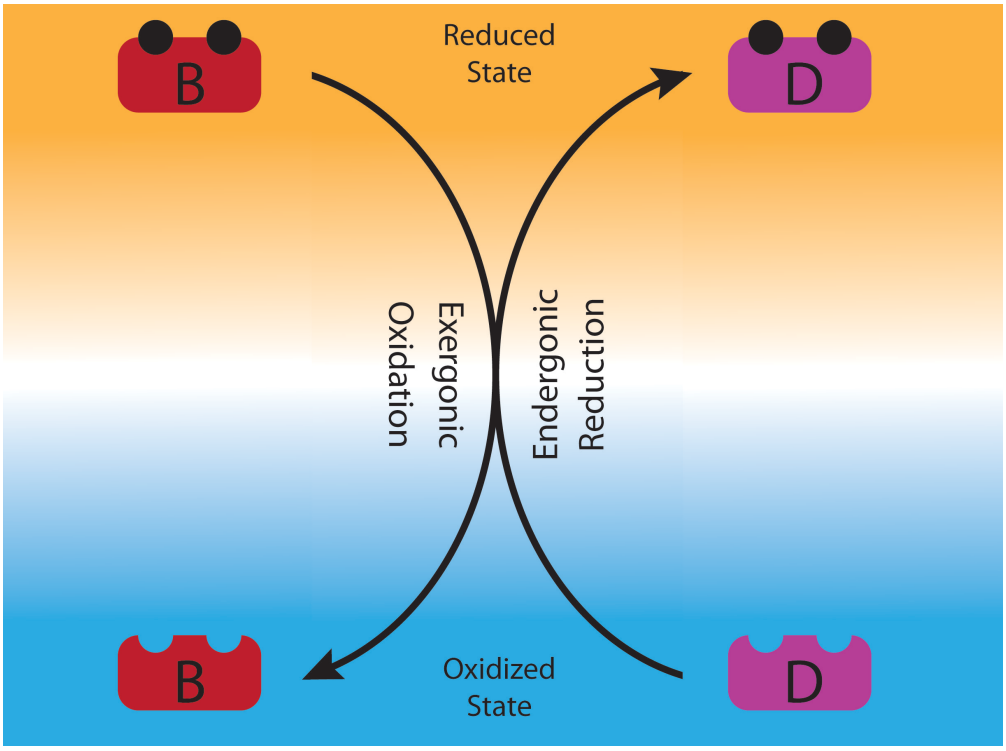


Figure 17.11 This image illustrates the coupling of oxidation and reduction reactions. Molecule B is being oxidized to a lower energy state while molecule D is being reduced to a higher energy state. (Image by Robert Bear)

Electron Carriers

In living systems, a small class of compounds functions as electron shuttles: They bind and carry high-energy electrons between compounds in pathways. The principal electron carriers we will consider are derived from the B vitamin group and are derivatives of nucleotides. These compounds can be easily reduced (that is, they accept electrons) or oxidized (they lose electrons). Nicotinamide adenine dinucleotide (NAD) (**Figure 17.12**) is derived from vitamin B3, niacin. NAD^+ is the oxidized form of the molecule; NADH is the reduced form of the molecule after it has accepted two electrons and a proton (which together are the equivalent of a hydrogen atom with an extra electron).

NAD^+ can accept electrons from an organic molecule according to the general equation:

$$\underset{\text{Reducing agent}}{RH} + \underset{\text{Oxidizing agent}}{NAD^+} \rightarrow \underset{\text{Reduced}}{NADH} + \underset{\text{Oxidized}}{R}$$

When electrons are added to a compound, they are reduced. A compound that reduces another is called a reducing agent. In the above equation, RH is a reducing agent, and NAD^+ is reduced to NADH. When electrons are removed from compound, it oxidized. A compound that oxidizes another is called an oxidizing agent. In the above equation, NAD^+ is an oxidizing agent, and RH is oxidized to R.

Similarly, flavin adenine dinucleotide (FAD^+) is derived from vitamin B_2, also called riboflavin. Its reduced form is $FADH_2$. A second variation of NAD, NADP, contains an extra phosphate group. Both NAD^+ and FAD^+ are extensively used in energy extraction from sugars, and NADP plays an important role in anabolic reactions and photosynthesis.

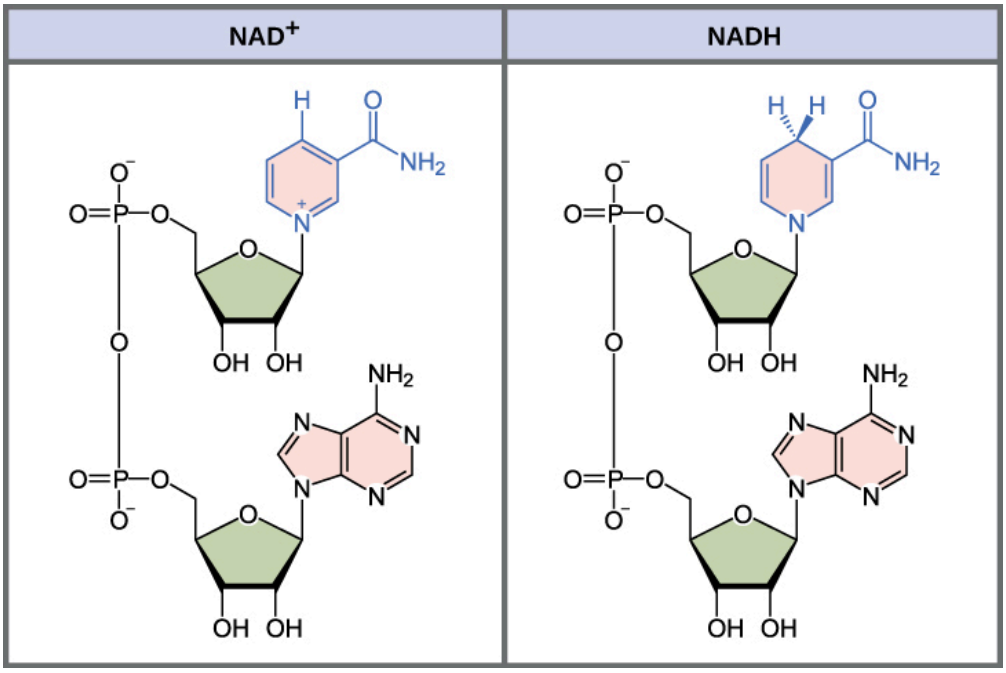


Figure 17.12 The oxidized form of the electron carrier (NAD^+) is shown on the left and the reduced form (NADH) is shown on the right. The nitrogenous base in NADH has one more hydrogen ion and two more electrons than in NAD^+.

Electron Transport Chain

The energy associated with high energy electrons is harvested in an incremental fashion, and electron transport chains allow for this type of energy harvesting. An **Electron Transport Chain** is a series of membrane bound proteins that are specialized in shuttling electrons (**Figure 17.13**). An electron donor drops off high energy electrons and as the electrons pass from one protein to the next in the chain, a small amount of energy is released. The energy release by each step is then available to do work. After the electrons pass through the transport chain, the electrons are removed from the chain and attached to the final electron acceptor. As you will see later in this module, electron transport chains play an important role in the processes of photosynthesis and cellular respiration.

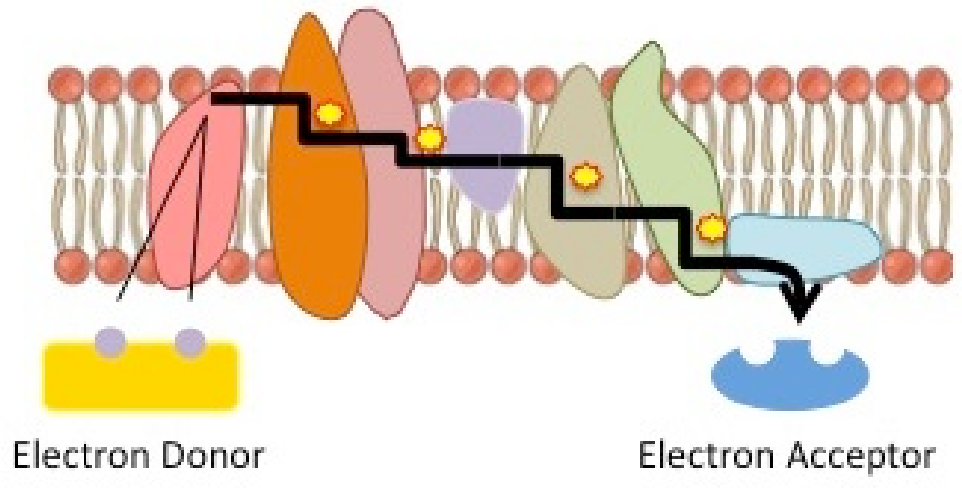


Figure 17.13 This image illustrates the release of energy (noted as yellow and red stars) that is available to do work as electrons pass from one protein to another in the electron transport chain. (Image by Robert Bear)

ATP in Living Systems

A living cell cannot store significant amounts of free energy. Excess free energy would result in an increase of heat in the cell, which would result in excessive thermal motion that could damage and then destroy the cell. Rather, a cell must be able to handle that energy in a way that enables the cell to store energy safely and release it for use only as needed. Living cells

accomplish this by using the compound adenosine triphosphate (ATP). ATP is often called the “energy currency” of the cell, and, like currency, this versatile compound can be used to fill any energy need of the cell. How? It functions similarly to a rechargeable battery.

When ATP is broken down, usually by the removal of its terminal phosphate group, energy is released. The energy is used to do work by the cell, usually by the released phosphate binding to another molecule, activating it. For example, in the mechanical work of muscle contraction, ATP supplies the energy to move the contractile muscle proteins. Recall the active transport work of the sodium-potassium pump in the biological membranes. ATP alters the structure of the integral protein that functions as the pump, changing its affinity for sodium and potassium. In this way, the cell performs work, pumping ions against their electrochemical gradients.

ATP Structure and Function

At the heart of ATP is a molecule of adenosine monophosphate (AMP), which is composed of an adenine molecule bonded to a ribose molecule and to a single phosphate group (**Figure 17.14**). Ribose is a five-carbon sugar found in RNA, and AMP is one of the nucleotides in RNA. The addition of a second phosphate group by a condensation reaction to this core molecule results in the formation of adenosine diphosphate (ADP); the addition of a third phosphate group by a condensation reaction forms adenosine triphosphate (ATP).

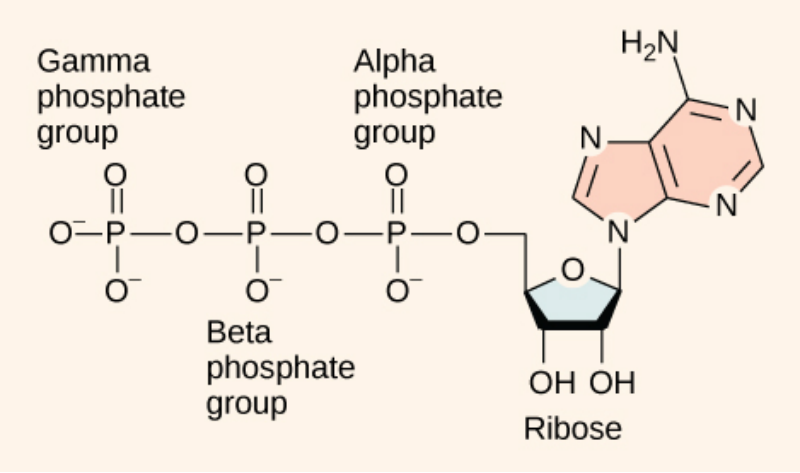


Figure 17.14 ATP (adenosine triphosphate) has three phosphate groups that can be removed by hydrolysis to form ADP (adenosine diphosphate) or AMP (adenosine monophosphate).The negative charges on the phosphate group naturally repel each other, requiring energy to bond them together and releasing energy when these bonds are broken.

The addition of a phosphate group to a molecule requires energy. Phosphate groups are negatively charged and thus repel one another when they are arranged in series, as they are in ADP and ATP. This repulsion makes the ADP and ATP molecules inherently unstable. The release of one or two phosphate groups from ATP, a process called **hydrolysis**, releases energy.

ATP Powers Cellular Work

Hydrolysis is the process of breaking complex macromolecules apart. During hydrolysis, water is split, or lysed, and the resulting hydrogen atom (H^+) and a hydroxyl group (OH^-) are added to the larger molecule. The hydrolysis of ATP produces ADP, together with an inorganic phosphate ion (P_i), and the release of free energy an exergonic reaction. This free energy is then available to drive an endergonic reaction. The release of energy and absorption of this free energy is a **coupled reaction**. To carry out life processes, ATP is continuously broken down into ADP, and like a rechargeable battery, ADP is continuously regenerated into ATP by the reattachment of a third phosphate group by a condensation reaction (**Figure 17.15**). Water, which was broken down into its hydrogen atom and hydroxyl group during ATP hydrolysis, is reformed when a third phosphate is added to the ADP molecule, reforming ATP with energy being absorbed. The energy necessary for driving the formation of ATP comes from an exergonic reaction. Once again the coupling of an exergonic reaction drives an endergonic reaction or the formation of ATP.

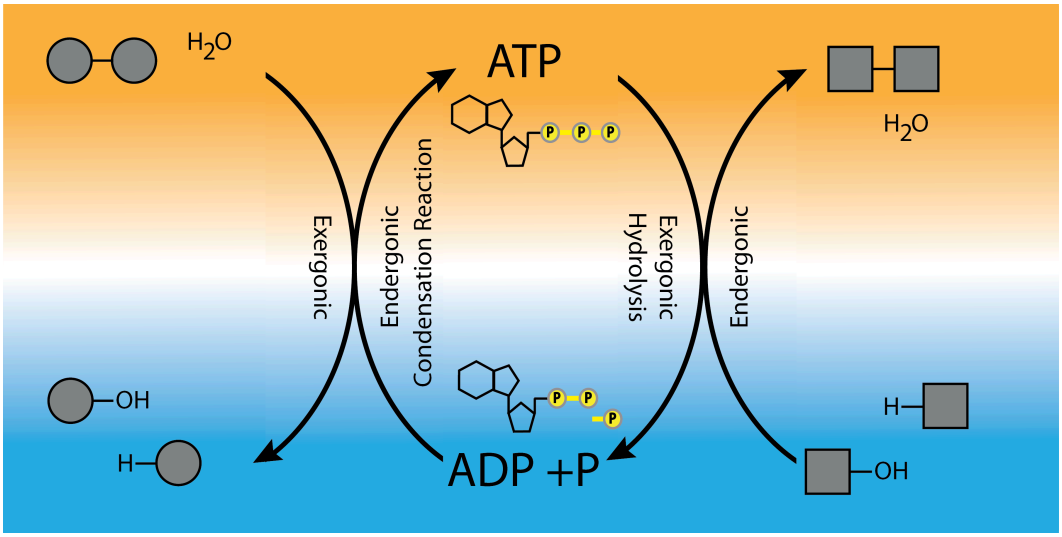


Figure 17.15 This image illustrates the ATP cycle and how ATP couples the release of energy from one reaction and makes that energy available for cellular work or endergonic reactions. (Image by Robert Bear)

Obviously, energy must be infused into the system to regenerate ATP. Where does this energy come from? In nearly every living thing on earth, the energy comes from the metabolism of glucose. In this way, ATP is a direct link between the limited set of exergonic pathways of glucose catabolism and the multitude of endergonic pathways that power living cells.

Phosphorylation

Recall that, in some chemical reactions, enzymes may bind to several substrates that react with each other on the enzyme, forming an intermediate complex. An intermediate complex is a temporary structure, and it allows one of the substrates (such as ATP) and reactants to more readily react with each other; in reactions involving ATP, ATP is one of the substrates and ADP is a product. During an endergonic chemical reaction, ATP forms an intermediate complex with the substrate and enzyme in the reaction. This intermediate complex allows the ATP to transfer its third phosphate group, with its energy, to the substrate, a process called phosphorylation. **Phosphorylation** refers to the addition of the phosphate (~P). This is illustrated by the following generic reaction:

$$\text{A} + \text{enzyme} + \text{ATP} \rightarrow [\text{A} - \text{enzyme} - \sim \text{P}] \rightarrow \text{B} + \text{enzyme} + \text{ADP} + \text{phosphate ion}$$

When the intermediate complex breaks apart, the energy is used to modify the substrate and convert it into a product of the reaction. The ADP molecule and a free phosphate ion are released into the medium and are available for recycling through cell metabolism.

Substrate Phosphorylation

ATP is generated through two mechanisms during the breakdown of glucose. A few ATP molecules are generated (that is, regenerated from ADP) as a direct result of the chemical reactions that occur in the catabolic pathways. A phosphate group is removed from an intermediate reactant in the pathway, and the free energy of the reaction is used to add the third phosphate to an available ADP molecule, producing ATP (**Figure 17.16**). This very direct method of phosphorylation is called **substrate-level phosphorylation**.

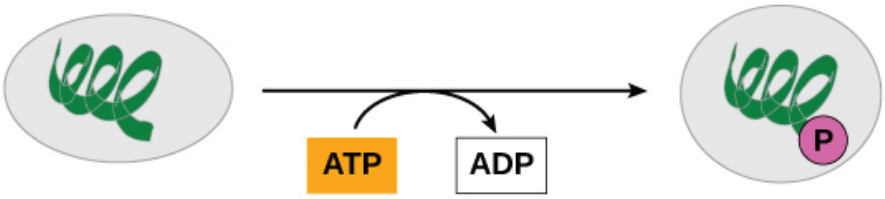


Figure 17.16 In phosphorylation reactions, the gamma phosphate of ATP is attached to a protein.

Oxidative Phosphorylation

Most of the ATP generated during glucose catabolism, however, is derived from a much more complex process, chemiosmosis, which takes place in mitochondria (**Figure 17.17**) within a eukaryotic cell or the plasma membrane of a

prokaryotic cell. **Chemiosmosis**, a process of ATP production in cellular metabolism, is used to generate 90 percent of the ATP made during glucose catabolism and is also the method used in the light reactions of photosynthesis to harness the energy of sunlight. The production of ATP using the process of chemiosmosis is called oxidative phosphorylation because of the involvement of oxygen in the process.

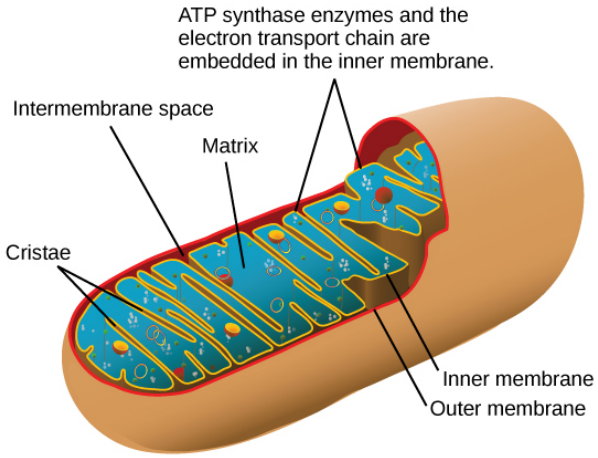


Figure 17.17 In eukaryotes, oxidative phosphorylation takes place in mitochondria. In prokaryotes, this process takes place in the plasma membrane. (Credit: modification of work by Mariana Ruiz Villareal)

17.5 | Cell Membranes and Passive Transport

Introduction

> “ A vital phenomenon can only be regarded as explained if it has been proven that it appears as the result of the material components of living organisms interacting according to the laws which those same components follow in their interactions outside of living systems. ”

A. E. Fick, German physiologist who developed the first mathematical explanation of diffusion

The simple diffusion model developed by Fick does indeed explain the movement of molecules in both living and non-living systems. Molecules move freely from an area of higher concentration to and area of lower concentration, according to Fick's Law. Living organisms, however, impose a barrier to diffusion of molecules into and from their cells - the plasma membrane which surrounds all cells. Plasma membranes must allow certain substances to enter and leave a cell, and prevent some harmful materials from entering and some essential materials from leaving. In other words, plasma membranes are **selectively permeable**—they allow some substances to pass through, but not others. If they were to lose this selectivity, the cell would no longer be able to sustain itself, and it would be destroyed. Some cells require larger amounts of specific substances than do other cells; they must have a way of obtaining these materials from extracellular fluids. This may happen passively, as certain materials move back and forth, or the cell may have special mechanisms that facilitate transport. Some materials are so important to a cell that it spends some of its energy, hydrolyzing adenosine triphosphate (ATP), to obtain these materials. Red blood cells use some of their energy doing just that. All cells spend the majority of their energy to maintain an imbalance of sodium and potassium ions between the interior and exterior of the cell.

The most direct forms of membrane transport are passive. **Passive transport** is a naturally occurring phenomenon and does not require the cell to exert any of its energy to accomplish the movement. In passive transport, substances move from an area of higher concentration to an area of lower concentration. A physical space in which there is a range of concentrations of a single substance is said to have a **concentration gradient**.

Selective Permeability

Plasma membranes are asymmetric: the interior of the membrane is not identical to the exterior of the membrane. In fact, there is a considerable difference between the array of phospholipids and proteins between the two leaflets that form a membrane. On the interior of the membrane, some proteins serve to anchor the membrane to fibers of the cytoskeleton. There are peripheral proteins on the exterior of the membrane that bind elements of the extracellular matrix. Carbohydrates, attached to lipids or proteins, are also found on the exterior surface of the plasma membrane. These carbohydrate complexes help the cell bind substances that the cell needs in the extracellular fluid. This adds considerably to the selective nature of plasma membranes (**Figure 17.18**).

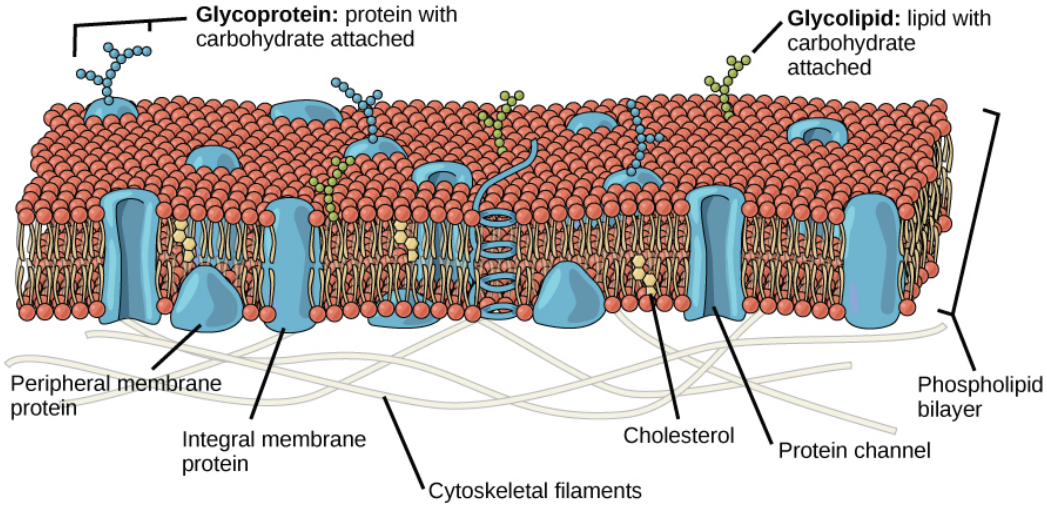


Figure 17.18 The exterior surface of the plasma membrane is not identical to the interior surface of the same membrane.

Recall that plasma membranes are amphiphilic: They have hydrophilic and hydrophobic regions. This characteristic helps the movement of some materials through the membrane and hinders the movement of others. Lipid-soluble material with a low molecular weight can easily slip through the hydrophobic lipid core of the membrane. Substances such as the fat-soluble vitamins A, D, E, and K readily pass through the plasma membranes in the digestive tract and other tissues. Fat-soluble drugs and hormones also gain easy entry into cells and are readily transported into the body's tissues and organs. Molecules of oxygen and carbon dioxide have no charge and so pass through membranes by simple diffusion.

Polar substances present problems for the membrane. While some polar molecules connect easily with the outside of a cell, they cannot readily pass through the lipid core of the plasma membrane. Additionally, while small ions could easily slip through the spaces in the mosaic of the membrane, their charge prevents them from doing so. Ions such as sodium, potassium, calcium, and chloride must have special means of penetrating plasma membranes. Simple sugars and amino acids also need help with transport across plasma membranes, achieved by various transmembrane proteins (channels).

Diffusion

Diffusion is a passive process of transport. A single substance tends to move from an area of high concentration to an area of low concentration until the concentration is equal across a space. You are familiar with diffusion of substances through the air. For example, think about someone opening a bottle of ammonia in a room filled with people. The ammonia gas is at its highest concentration in the bottle; its lowest concentration is at the edges of the room. The ammonia vapor will diffuse, or spread away, from the bottle, and gradually, more and more people will smell the ammonia as it spreads. Materials move within the cell's cytosol by diffusion, and certain materials move through the plasma membrane by diffusion (**Figure 17.19**). Diffusion expends no energy. On the contrary, concentration gradients are a form of potential energy, dissipated as the gradient is eliminated.

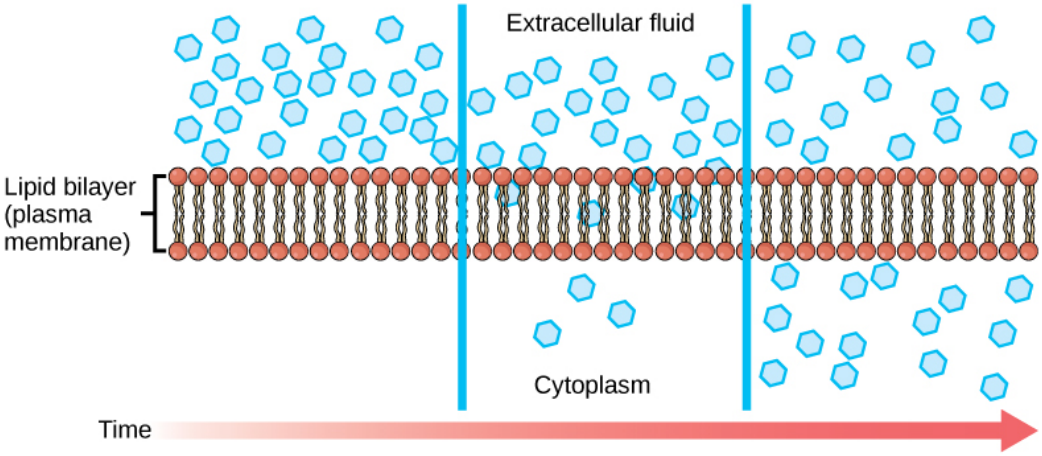


Figure 17.19 Diffusion through a permeable membrane moves a substance from an area of high concentration (extracellular fluid, in this case) down its concentration gradient (into the cytoplasm). (credit: modification of work by Mariana Ruiz Villareal)

Each separate substance in a medium, such as the extracellular fluid, has its own concentration gradient, independent of the concentration gradients of other materials. In addition, each substance will diffuse according to that gradient. Within a system, there will be different rates of diffusion of the different substances in the medium.

Factors That Affect Diffusion

Molecules move constantly in a random manner, at a rate that depends on their mass, their environment, and the amount of thermal energy they possess, which in turn is a function of temperature. This movement accounts for the diffusion of molecules through whatever medium in which they are localized. A substance will tend to move into any space available to it until it is evenly distributed throughout it. After a substance has diffused completely through a space, removing its concentration gradient, molecules will still move around in the space, but there will be no *net* movement of the number of molecules from one area to another. This lack of a concentration gradient in which there is no net movement of a substance is known as dynamic equilibrium. While diffusion will go forward in the presence of a concentration gradient of a substance, several factors affect the rate of diffusion.

- Extent of the concentration gradient: The greater the difference in concentration, the more rapid the diffusion. The closer the distribution of the material gets to equilibrium, the slower the rate of diffusion becomes.
- Mass of the molecules diffusing: Heavier molecules move more slowly; therefore, they diffuse more slowly. The reverse is true for lighter molecules.
- Temperature: Higher temperatures increase the energy and therefore the movement of the molecules, increasing the rate of diffusion. Lower temperatures decrease the energy of the molecules, thus decreasing the rate of diffusion.
- Solvent density: As the density of a solvent increases, the rate of diffusion decreases. The molecules slow down because they have a more difficult time getting through the denser medium. If the medium is less dense, diffusion increases. Because cells primarily use diffusion to move materials within the cytoplasm, any increase in the cytoplasm's density will inhibit the movement of the materials. An example of this is a person experiencing dehydration. As the body's cells lose water, the rate of diffusion decreases in the cytoplasm, and the cells' functions deteriorate. Neurons tend to be very sensitive to this effect. Dehydration frequently leads to unconsciousness and possibly coma because of the decrease in diffusion rate within the cells.
- Solubility: As discussed earlier, nonpolar or lipid-soluble materials pass through plasma membranes more easily than polar materials, allowing a faster rate of diffusion.
- Surface area and thickness of the plasma membrane: Increased surface area increases the rate of diffusion, whereas a thicker membrane reduces it.
- Distance travelled: The greater the distance that a substance must travel, the slower the rate of diffusion. This places an upper limitation on cell size. A large, spherical cell will die because nutrients or waste cannot reach or leave the center of the cell, respectively. Therefore, cells must either be small in size, as in the case of many prokaryotes, or be flattened, as with many single-celled eukaryotes.

A variation of diffusion is the process of filtration. In filtration, material moves according to its concentration gradient through a membrane; sometimes the rate of diffusion is enhanced by pressure, causing the substances to filter more rapidly.

This occurs in the kidney, where blood pressure forces large amounts of water and accompanying dissolved substances, or **solutes**, out of the blood and into the renal tubules. The rate of diffusion in this instance is almost totally dependent on pressure. One of the effects of high blood pressure is the appearance of protein in the urine, which is "squeezed through" by the abnormally high pressure.

Facilitated diffusion

In **facilitated diffusion**, also called facilitated transport, materials diffuse across the plasma membrane with the help of membrane proteins. A concentration gradient exists that would allow these materials to diffuse into the cell without expending cellular energy. However, these materials are ions or polar molecules that are repelled by the hydrophobic parts of the biological membrane. Facilitated transport proteins shield these materials from the repulsive force of the membrane, allowing them to diffuse into the cell.

The material being transported is first attached to protein or glycoprotein receptors on the exterior surface of the plasma membrane. This allows the material that is needed by the cell to be removed from the extracellular fluid. The substances are then passed to specific integral proteins that facilitate their passage. Some of these integral proteins are collections of beta pleated sheets that form a pore or channel through the phospholipid bilayer. Others are carrier proteins which bind with the substance and aid its diffusion through the membrane.

Channels

The integral proteins involved in facilitated transport are collectively referred to as **transport proteins**, and they function as either channels for the material or carriers. In both cases, they are transmembrane proteins. Channels are specific for the substance that is being transported. **Channel proteins** have hydrophilic domains exposed to the intracellular and extracellular fluids; they additionally have a hydrophilic channel through their core that provides a hydrated opening through the membrane layers (**Figure 17.20**). Passage through the channel allows polar compounds to avoid the nonpolar central layer of the plasma membrane that would otherwise slow or prevent their entry into the cell. **Aquaporins** are channel proteins that allow water to pass through the membrane at a very high rate.

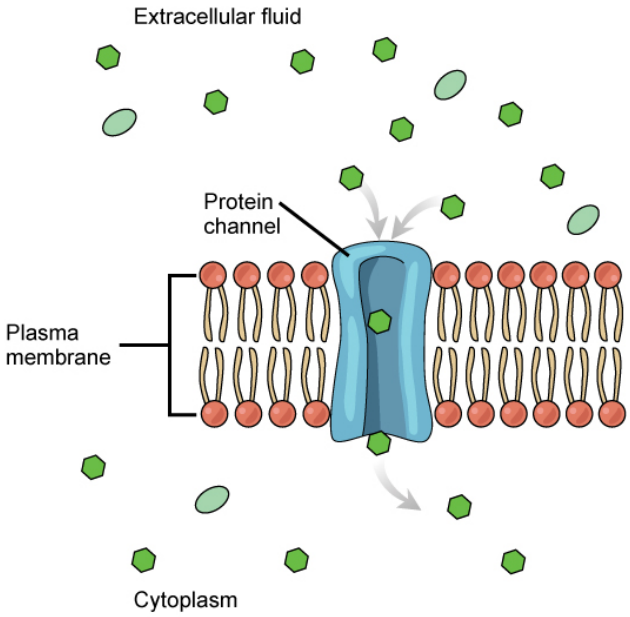


Figure 17.20 Facilitated transport moves substances down their concentration gradients. They may cross the plasma membrane with the aid of channel proteins. (credit: modification of work by Mariana Ruiz Villareal)

Channel proteins are either open at all times or they are "gated," which controls the opening of the channel. The attachment of a particular ion to the channel protein may control the opening, or other mechanisms or substances may be involved. In some tissues, sodium and chloride ions pass freely through open channels, whereas in other tissues a gate must be opened to allow passage. An example of this occurs in the kidney, where both forms of channels are found in different parts of the renal tubules. Cells involved in the transmission of electrical impulses, such as nerve and muscle cells, have gated channels for sodium, potassium, and calcium in their membranes. Opening and closing of these channels changes the relative concentrations on opposing sides of the membrane of these ions, resulting in the facilitation of electrical transmission along membranes (in the case of nerve cells) or in muscle contraction (in the case of muscle cells).

Carrier Proteins

Another type of protein embedded in the plasma membrane is a **carrier protein**. This aptly named protein binds a substance and, in doing so, triggers a change of its own shape, moving the bound molecule from the outside of the cell to its interior (Figure 17.21); depending on the gradient, the material may move in the opposite direction. Carrier proteins are typically specific for a single substance. This selectivity adds to the overall selectivity of the plasma membrane. The exact mechanism for the change of shape is poorly understood. Proteins can change shape when their hydrogen bonds are affected, but this may not fully explain this mechanism. Each carrier protein is specific to one substance, and there are a finite number of these proteins in any membrane. This can cause problems in transporting enough of the material for the cell to function properly. When all of the proteins are bound to their ligands, they are saturated and the rate of transport is at its maximum. Increasing the concentration gradient at this point will not result in an increased rate of transport.

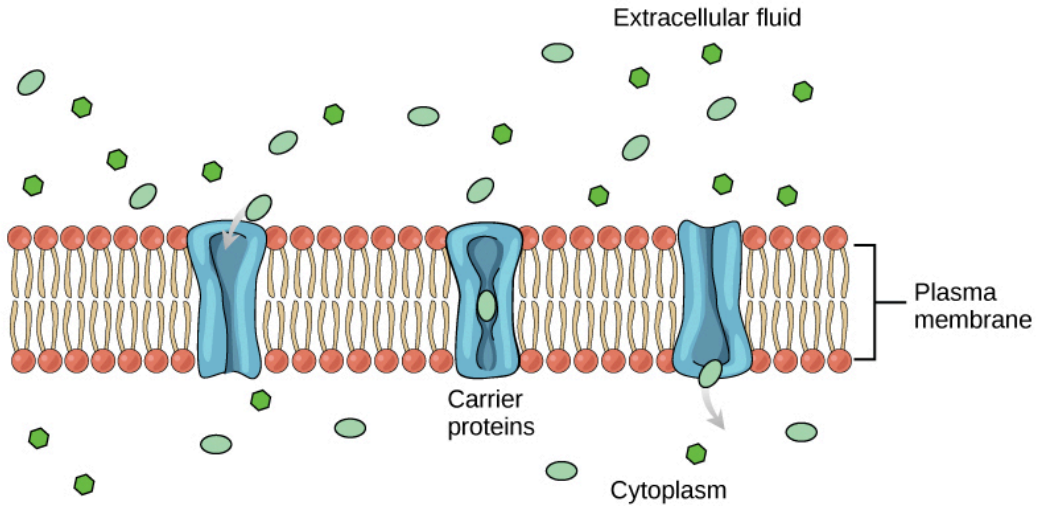


Figure 17.21 Some substances are able to move down their concentration gradient across the plasma membrane with the aid of carrier proteins. Carrier proteins change shape as they move molecules across the membrane. (credit: modification of work by Mariana Ruiz Villareal)

An example of this process occurs in the kidney. Glucose, water, salts, ions, and amino acids needed by the body are filtered in one part of the kidney. This filtrate, which includes glucose, is then reabsorbed in another part of the kidney. Because there are only a finite number of carrier proteins for glucose, if more glucose is present than the proteins can handle, the excess is not transported and it is excreted from the body in the urine. In a diabetic individual, this is described as “spilling glucose into the urine.” A different group of carrier proteins called glucose transport proteins, or GLUTs, are involved in transporting glucose and other hexose sugars through plasma membranes within the body.

Channel and carrier proteins transport material at different rates. Channel proteins transport much more quickly than do carrier proteins. Channel proteins facilitate diffusion at a rate of tens of millions of molecules per second, whereas carrier proteins work at a rate of a thousand to a million molecules per second.

Osmosis

Osmosis is the movement of water through a semipermeable membrane according to the concentration gradient of water across the membrane, which is inversely proportional to the concentration of solutes. While diffusion transports material across membranes and within cells, osmosis transports *only water* across a membrane and the membrane limits the diffusion of solutes in the water. Not surprisingly, the aquaporins that facilitate water movement play a large role in osmosis, most prominently in red blood cells and the membranes of kidney tubules.

Mechanism

Osmosis is a special case of diffusion. Water, like other substances, moves from an area of high concentration to one of low concentration. An obvious question is what makes water move at all? Imagine a beaker with a semipermeable membrane separating the two sides or halves (Figure 17.22). On both sides of the membrane the water level is the same, but there are different concentrations of a dissolved substance, or **solute**, that cannot cross the membrane (otherwise the concentrations on each side would be balanced by the solute crossing the membrane). If the volume of the solution on both sides of the membrane is the same, but the concentrations of solute are different, then there are different amounts of water, the solvent, on either side of the membrane.

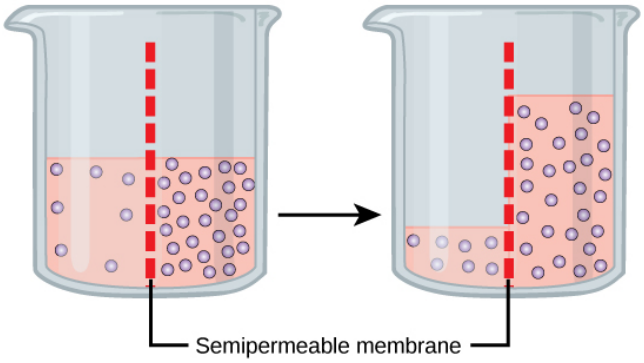


Figure 17.22 In osmosis, water always moves from an area of higher water concentration to one of lower concentration. In the diagram shown, the solute cannot pass through the selectively permeable membrane, but the water can.

To illustrate this, imagine two full glasses of water. One has a single teaspoon of sugar in it, whereas the second one contains one-quarter cup of sugar. If the total volume of the solutions in both cups is the same, which cup contains more water? Because the large amount of sugar in the second cup takes up much more space than the teaspoon of sugar in the first cup, the first cup has more water in it.

Returning to the beaker example, recall that it has a mixture of solutes on either side of the membrane. A principle of diffusion is that the molecules move around and will spread evenly throughout the medium if they can. However, only the material capable of getting through the membrane will diffuse through it. In this example, the solute cannot diffuse through the membrane, but the water can. Water has a concentration gradient in this system. Thus, water will diffuse down its concentration gradient, crossing the membrane to the side where it is less concentrated. This diffusion of water through the membrane—osmosis—will continue until the concentration gradient of water goes to zero or until the hydrostatic pressure of the water balances the osmotic pressure. Osmosis proceeds constantly in living systems.

Tonicity

Tonicity is a measure of the osmotic pressure of two solutions separated by a semipermeable membrane. So, tonicity describes how an extracellular solution can change the volume of a cell by affecting osmosis. A solution's tonicity often directly correlates with the osmolarity of the solution. **Osmolarity** describes the total solute concentration of the solution. A solution with low osmolarity has a greater number of water molecules relative to the number of solute particles; a solution with high osmolarity has fewer water molecules with respect to solute particles. In a situation in which solutions of two different osmolarities are separated by a membrane permeable to water, though not to the solute, water will move from the side of the membrane with lower osmolarity (and more water) to the side with higher osmolarity (and less water). This effect makes sense if you remember that the solute cannot move across the membrane, and thus the only component in the system that can move—the water—moves along its own concentration gradient. An important distinction that concerns living systems is that osmolarity measures the number of particles (which may be molecules) in a solution. Therefore, a solution that is cloudy with cells may have a lower osmolarity than a solution that is clear, if the second solution contains more dissolved molecules than there are cells.

Hypotonic Solutions

Three terms—hypotonic, isotonic, and hypertonic—are used to relate the osmolarity of a cell to the osmolarity of the extracellular fluid that contains the cells. In a **hypotonic** situation, the extracellular fluid has lower osmolarity than the fluid inside the cell, and water enters the cell. (In living systems, the point of reference is always the cytoplasm, so the prefix *hypo-* means that the extracellular fluid has a lower concentration of solutes, or a lower osmolarity, than the cell cytoplasm.) It also means that the extracellular fluid has a higher concentration of water in the solution than does the cell. In this situation, water will follow its concentration gradient and enter the cell.

Hypertonic Solutions

As for a **hypertonic** solution, the prefix *hyper-* refers to the extracellular fluid having a higher osmolarity than the cell's cytoplasm; therefore, the fluid contains less water than the cell does. Because the cell has a relatively higher concentration of water, water will leave the cell.

Isotonic Solutions

In an **isotonic** solution, the extracellular fluid has the same osmolarity as the cell. If the osmolarity of the cell matches that of the extracellular fluid, there will be no net movement of water into or out of the cell, although water will still move in and

out. Blood cells and plant cells in hypertonic, isotonic, and hypotonic solutions take on characteristic appearances (**Figure 17.23**).

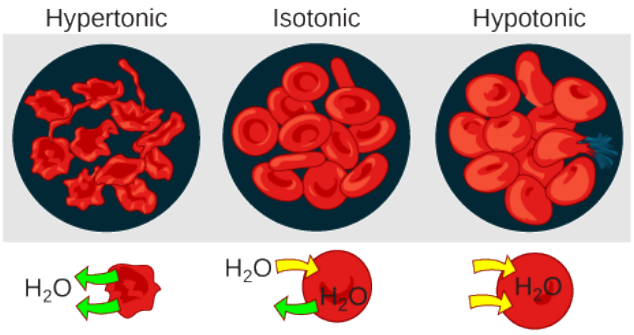


Figure 17.23 Osmotic pressure changes the shape of red blood cells in hypertonic, isotonic, and hypotonic solutions. (credit: Mariana Ruiz Villareal)

Tonicity in Living Systems

In a hypotonic environment, water enters a cell, and the cell swells. In an isotonic condition, the relative concentrations of solute and solvent are equal on both sides of the membrane. There is no net water movement; therefore, there is no change in the size of the cell. In a hypertonic solution, water leaves a cell and the cell shrinks. If either the hypo- or hyper- condition goes to excess, the cell's functions become compromised, and the cell may be destroyed.

A red blood cell will burst, or lyse, when it swells beyond the plasma membrane's capability to expand. Remember, the membrane resembles a mosaic, with discrete spaces between the molecules composing it. If the cell swells, and the spaces between the lipids and proteins become too large, the cell will break apart.

In contrast, when excessive amounts of water leave a red blood cell, the cell shrinks, or crenates. This has the effect of concentrating the solutes left in the cell, making the cytosol denser and interfering with diffusion within the cell. The cell's ability to function will be compromised and may also result in the death of the cell.

Various living things have ways of controlling the effects of osmosis—a mechanism called osmoregulation. Some organisms, such as plants, fungi, bacteria, and some protists, have cell walls that surround the plasma membrane and prevent cell lysis in a hypotonic solution. The plasma membrane can only expand to the limit of the cell wall, so the cell will not lyse. In fact, the cytoplasm in plants is always slightly hypertonic to the cellular environment, and water will always enter a cell if water is available. This inflow of water produces turgor pressure, which stiffens the cell walls of the plant (**Figure 17.24**). In nonwoody plants, turgor pressure supports the plant. Conversly, if the plant is not watered, the extracellular fluid will become hypertonic, causing water to leave the cell. In this condition, the cell does not shrink because the cell wall is not flexible. However, the plasma membrane detaches from the wall and constricts the cytoplasm. This is called **plasmolysis**. Plants lose turgor pressure in this condition and wilt (**Figure 17.25**).

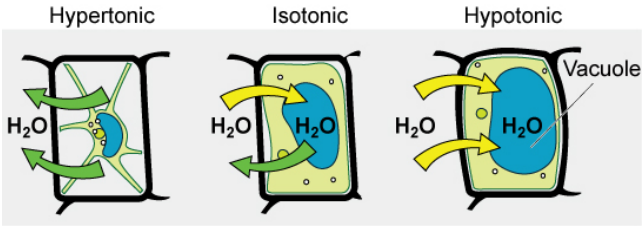


Figure 17.24 The turgor pressure within a plant cell depends on the tonicity of the solution that it is bathed in. (credit: modification of work by Mariana Ruiz Villareal)

Figure 17.25 Without adequate water, the plant on the left has lost turgor pressure, visible in its wilting; the turgor pressure is restored by watering it (right). (credit: Victor M. Vicente Selvas)

Tonicity is a concern for all living things. For example, paramecia and amoebas, which are protists that lack cell walls, have contractile vacuoles. This vesicle collects excess water from the cell and pumps it out, keeping the cell from lysing as it takes on water from its environment (**Figure 17.26**).

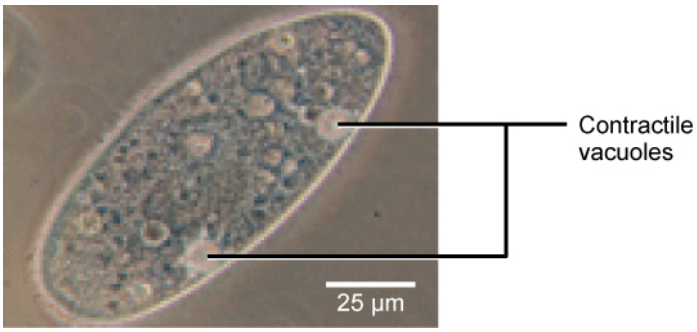


Figure 17.26 A paramecium's contractile vacuole, here visualized using bright field light microscopy at 480x magnification, continuously pumps water out of the organism's body to keep it from bursting in a hypotonic medium. (credit: modification of work by NIH; scale-bar data from Matt Russell)

Many marine invertebrates have internal salt levels matched to their environments, making them isotonic with the water in which they live. Fish, however, must spend approximately five percent of their metabolic energy maintaining osmotic homeostasis. Freshwater fish live in an environment that is hypotonic to their cells. These fish actively take in salt through their gills and excrete diluted urine to rid themselves of excess water. Saltwater fish live in the reverse environment, which is hypertonic to their cells, and they secrete salt through their gills and excrete highly concentrated urine.

In vertebrates, the kidneys regulate the amount of water in the body. Osmoreceptors are specialized cells in the brain that monitor the concentration of solutes in the blood. If the levels of solutes increase beyond a certain range, a hormone is released that retards water loss through the kidney and dilutes the blood to safer levels. Animals also have high concentrations of albumin, which is produced by the liver, in their blood. This protein is too large to pass easily through plasma membranes and is a major factor in controlling the osmotic pressures applied to tissues.

17.6 | Energy Requiring Transport

Introduction

"Against the wind,"

> “We were runnin' against the wind,”
>
> “We were young and strong and we were runnin'”
>
> “against the wind”

- Bob Seger, chorus of *Against the Wind.*

Active Transport

As the song says, you have to be strong to be moving against the wind. And, at the cellular level, you have to have significant energy to move materials against a concentration gradient. The process which cells use to do that is called active transport. **Active transport** mechanisms require the use of the cell’s energy, usually in the form of adenosine triphosphate (ATP). If a substance must move into the cell against its concentration gradient, that is, if the concentration of the substance inside the cell must be greater than its concentration in the extracellular fluid, the cell must use energy to move the substance. Some active transport mechanisms move small-molecular weight material, such as ions, through the membrane.

In addition to moving small ions and molecules through the membrane, cells also need to remove and take in larger molecules and particles. Some cells are even capable of engulfing entire unicellular microorganisms. You might have correctly hypothesized that the uptake and release of large particles by the cell requires energy. A large particle, however, cannot pass through the membrane, even with energy supplied by the cell.

Electrochemical Gradient

We have discussed simple concentration gradients—differential concentrations of a substance across a space or a membrane—but in living systems, gradients are more complex. Because cells contain proteins, most of which are negatively charged, and because ions move into and out of cells, there is an electrical gradient, a difference of charge, across the plasma membrane. The interior of living cells is electrically negative with respect to the extracellular fluid in which they are bathed; at the same time, cells have higher concentrations of potassium (K^+) and lower concentrations of sodium (Na^+) than does the extracellular fluid. Thus, in a living cell, the concentration gradient and electrical gradient of Na^+ promotes diffusion of the ion into the cell, and the electrical gradient of Na^+ (a positive ion) tends to drive it inward to the negatively charged interior. The situation is more complex, however, for other elements such as potassium. The electrical gradient of K^+ promotes diffusion of the ion *into* the cell, but the concentration gradient of K^+ promotes diffusion *out* of the cell (**Figure 17.27**). The combined gradient that affects an ion is called its **electrochemical gradient**, and it is especially important to muscle and nerve cells.

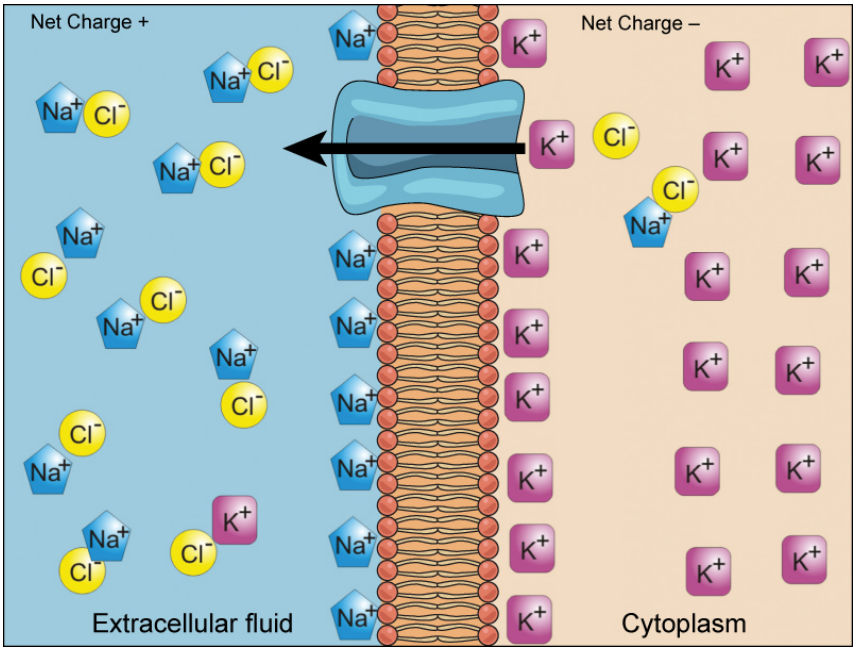


Figure 17.27 Electrochemical gradients arise from the combined effects of concentration gradients and electrical gradients. (credit: modification of work by “Synaptitude”/Wikimedia Commons)

Moving Against a Gradient

To move substances against a concentration or an electrochemical gradient, the cell must use energy. This energy is harvested from ATP that is generated through cellular metabolism. Active transport mechanisms, collectively called pumps or carrier proteins, work against electrochemical gradients. With the exception of ions, small substances constantly pass through plasma membranes. Active transport maintains concentrations of ions and other substances needed by living cells in the face of these passive changes. Much of a cell’s supply of metabolic energy may be spent maintaining these processes. Because active transport mechanisms depend on cellular metabolism for energy, they are sensitive to many metabolic poisons that interfere with the supply of ATP.

Two mechanisms exist for the transport of small-molecular weight material and macromolecules. Primary active transport moves ions across a membrane and creates a difference in charge across that membrane. The primary active transport system uses ATP to move a substance, such as an ion, into the cell, and often at the same time, a second substance is moved out of the cell. The sodium-potassium pump, an important pump in animal cells, expends energy to move potassium ions into the cell and a different number of sodium ions out of the cell (**Figure 17.28**). The action of this pump results in a concentration and charge difference across the membrane.

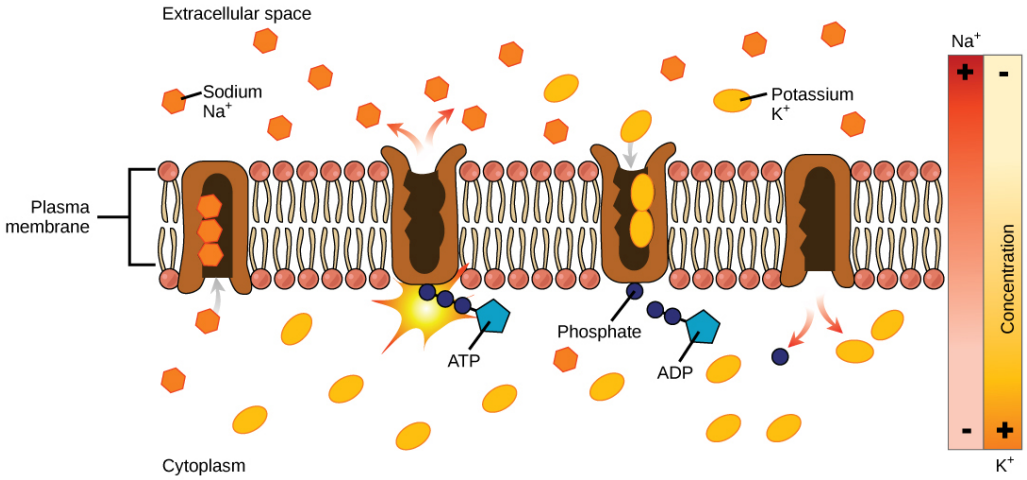


Figure 17.28 The sodium-potassium pump move potassium and sodium ions across the plasma membrane. (credit: modification of work by Mariana Ruiz Villarreal)

Secondary active transport describes the movement of material using the energy of the electrochemical gradient established by primary active transport. Using the energy of the electrochemical gradient created by the primary active transport system, other substances such as amino acids and glucose can be brought into the cell through membrane channels. ATP itself is formed through secondary active transport using a hydrogen ion gradient in the mitochondrion.

Bulk Transport of Materials

Endocytosis

Endocytosis is a type of energy requiring transport that moves particles, such as large molecules, parts of cells, and even whole cells, into a cell. There are different variations of endocytosis, but all share a common characteristic: The plasma membrane of the cell invaginates, forming a pocket around the target particle. The pocket pinches off, resulting in the particle being contained in a newly created vacuole that is formed from the plasma membrane.

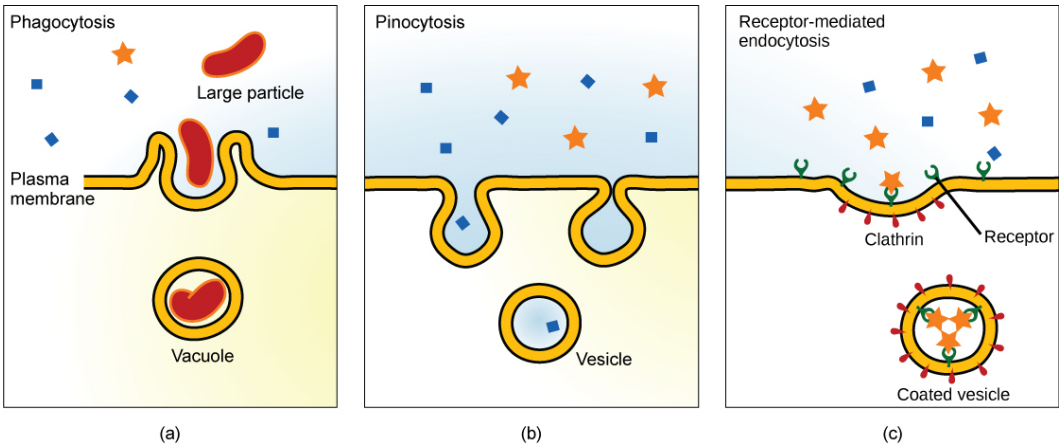


Figure 17.29 Three variations of endocytosis are shown. (a) In one form of endocytosis, phagocytosis, the plasma membrane surrounds the particle and pinches off to form an intracellular vacuole. (b) In another type of endocytosis, pinocytosis, the cell membrane surrounds a small volume of fluid and pinches off, forming a vesicle. (c) In receptor-mediated endocytosis, uptake of substances by the cell is targeted to a single type of substance that binds at the receptor on the plasma membrane. (credit: modification of work by Mariana Ruiz Villarreal)

Phagocytosis is the process by which large particles, such as cells, are taken in by a cell. For example, when microorganisms invade the human body, a type of white blood cell called a neutrophil removes the invader through this process, surrounding and engulfing the microorganism, which is then destroyed by the neutrophil (**Figure 17.29**).

A variation of endocytosis is called **pinocytosis**. This literally means “cell drinking” and was named at a time when the assumption was that the cell was purposefully taking in extracellular fluid. In reality, this process takes in solutes that the cell needs from the extracellular fluid (**Figure 17.29**).

A targeted variation of endocytosis employs binding proteins in the plasma membrane that are specific for certain substances (**Figure 17.29**). The particles bind to the proteins and the plasma membrane invaginates, bringing the substance and the proteins into the cell. If passage across the membrane of the target of **receptor-mediated endocytosis** is ineffective, it will not be removed from the tissue fluids or blood. Instead, it will stay in those fluids and increase in concentration. Some human diseases are caused by a failure of receptor-mediated endocytosis. For example, the form of cholesterol termed low-density lipoprotein or LDL (also referred to as “bad” cholesterol) is removed from the blood by receptor-mediated endocytosis. In the human genetic disease familial hypercholesterolemia, the LDL receptors are defective or missing entirely. People with this condition have life-threatening levels of cholesterol in their blood, because their cells cannot clear the chemical from their blood.

Exocytosis

In contrast to these methods of moving material into a cell is the process of exocytosis. **Exocytosis** is the opposite of the processes discussed above in that its purpose is to expel material from the cell into the extracellular fluid. A particle enveloped in membrane fuses with the interior of the plasma membrane. This fusion opens the membranous envelope to the exterior of the cell, and the particle is expelled into the extracellular space (**Figure 17.30**).

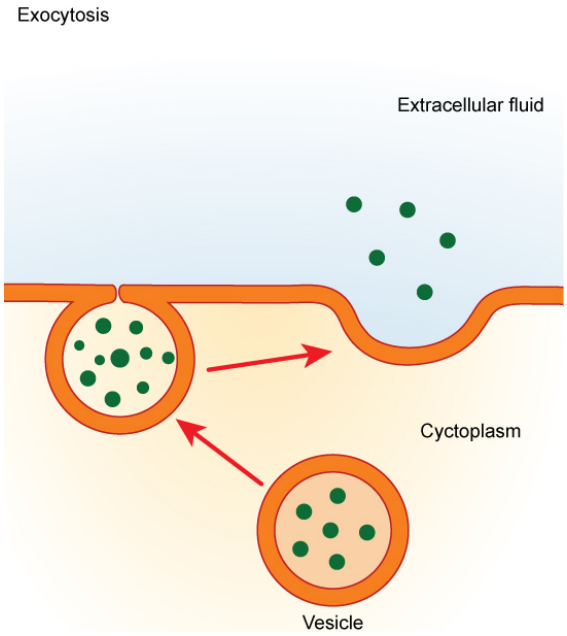


Figure 17.30 In exocytosis, a vesicle migrates to the plasma membrane, binds, and releases its contents to the outside of the cell. (credit: modification of work by Mariana Ruiz Villarreal)

18 | PHOTOSYNTHESIS

18.1 | Overview of Photosynthesis

Introduction

> “Nature has put itself the problem how to catch in flight light streaming to the earth and to store the most elusive of all powers in rigid form. To achieve this aim, it has covered the crust of earth with organisms which in their life processes absorb the light of the sun and use this power to produce a continuously accumulating chemical difference. ... The plants take in one form of power, light; and produce another power, chemical difference. ”

Robert Mayer, German physicist who developed the concept of the conservation of heat, 1845

The sun not only provides the energy for the plants themselves, but for all other creatures on the planet. Indeed, ancient photosynthesis provided the fossil fuel energy that we use to generate electricity or power our cars. Each cell in every organism runs on the chemical energy found mainly in carbohydrate molecules (food), and the majority of these molecules are produced by one process: photosynthesis. Through photosynthesis, certain organisms convert solar energy (sunlight) into chemical energy, which is then used to build carbohydrate molecules. The energy used to hold these molecules together is released when an organism breaks down food during cellular respiration. Cells then use this energy to perform work, such as movement.

The energy that is harnessed from photosynthesis enters the ecosystems of our planet continuously and is transferred from one organism to another until almost all of the harvested energy is transferred and released as heat energy. Therefore, directly or indirectly, the process of photosynthesis provides most of the energy required by living things on earth (**Figure 18.1**).

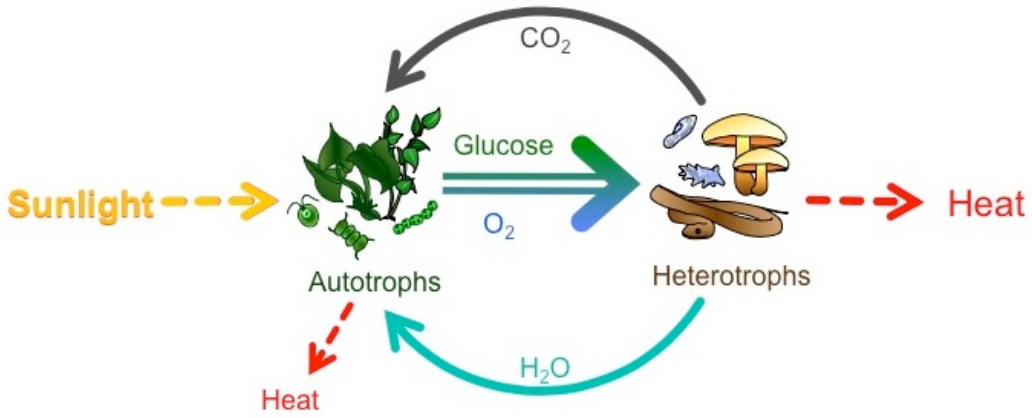


Figure 18.1 This image shows the role of photosynthesis is the flow of energy through and ecosystem. Light energy enters the system through photosynthesis and leaves primarily as heat energy once the energy is used by organisms. (image by Eva Horne and Robert Bear)

Photosynthesis also results in the release of oxygen into the atmosphere. In short, to eat and breathe, humans depend almost entirely on the organisms that carry out photosynthesis.

Solar Dependence and Food Production

Some organisms can carry out photosynthesis, whereas others cannot. An **autotroph** is an organism that can produce its own food. The Greek roots of the word *autotroph* mean "self" (*auto*) "feeder" (*troph*). Plants are the best-known autotrophs, but others exist, including certain types of bacteria and algae (**Figure 18.2**). Oceanic algae contribute enormous quantities of food and oxygen to global food chains. Plants are also **photoautotrophs**, a type of autotroph that uses sunlight and carbon from carbon dioxide to synthesize chemical energy in the form of carbohydrates. All organisms carrying out photosynthesis require sunlight.

(a) (b) (c)

Figure 18.2 (a) Plants, (b) algae, and (c) certain bacteria, called cyanobacteria, are photoautotrophs that can carry out photosynthesis. Algae can grow over enormous areas in water, at times completely covering the surface. (credit a: Steve Hillebrand, U.S. Fish and Wildlife Service; credit b: "eutrophication&hypoxia"/Flickr; credit c: NASA; scale-bar data from Matt Russell)

Heterotrophs are organisms incapable of photosynthesis that must therefore obtain energy and carbon from food by consuming other organisms. The Greek roots of the word *heterotroph* mean "other" (*hetero*) "feeder" (*troph*), meaning that their food comes from other organisms. Even if the food organism is another animal, this food traces its origins back to autotrophs and the process of photosynthesis. Humans are heterotrophs, as are all animals. Heterotrophs depend on autotrophs, either directly or indirectly. Deer and wolves are heterotrophs. A deer obtains energy by eating plants. A wolf eating a deer obtains energy that originally came from the plants eaten by that deer. The energy in the plant came from photosynthesis, and therefore it is the only autotroph in this example. Using this reasoning, all food eaten by humans also links back to autotrophs that carry out photosynthesis.

The Flow of Energy

Whether the organism is a bacterium, plant, or animal, all living things access energy by breaking down carbohydrate molecules. But if plants make carbohydrate molecules, why would they need to break them down, especially when it has been shown that the gas organisms release as a "waste product" (CO_2) acts as a substrate for the formation of more food in photosynthesis? Remember, living things need energy to perform life functions. In addition, an organism can either make its own food or eat another organism—either way, the food still needs to be broken down. Finally, in the process of breaking down food, called cellular respiration, heterotrophs release needed energy and produce "waste" in the form of CO_2 gas.

In nature, there is no such thing as waste. Every single atom of matter and energy is conserved, recycling over and over infinitely. Substances change form or move from one type of molecule to another, but their constituent atoms never disappear.

CO_2 is no more a form of waste than oxygen is form of waste from photosynthesis. Both are byproducts of reactions and move on to other reactions. Photosynthesis absorbs light energy to build carbohydrates in chloroplasts, and aerobic cellular respiration releases energy by using oxygen to take metabolize carbohydrates in the cytoplasm and mitochondria(**Figure 18.3**). Both processes use electron transport chains to capture the energy necessary to drive the reactions, because breaking down a substance requires energy. These two powerhouse processes, photosynthesis and cellular respiration, function in biological, cyclical harmony to allow organisms to access life-sustaining energy that originates millions of miles away in a burning star humans call the sun.

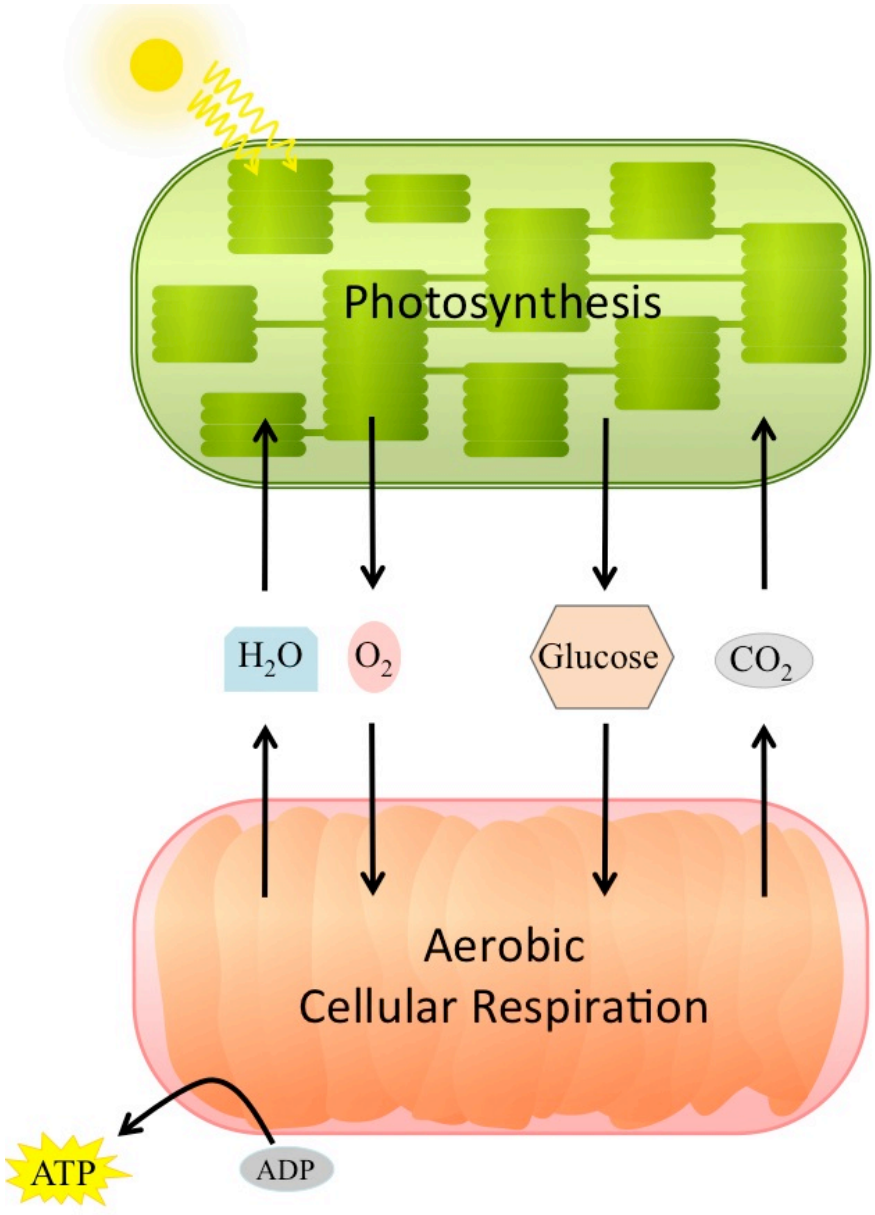


Figure 18.3 Photosynthesis which occurs in the chloroplast consumes carbon dioxide and water while producing carbohydrates (glucose) and oxygen while Aerobic Cellular respiration which occurs in the mitochondria consumes glucose and oxygen while producing carbohydrates. (Image by Eva Horne and Robert Bear)

Main Structures and Summary of Photosynthesis

Photosynthesis requires sunlight, carbon dioxide, and water as starting reactants (**Figure 18.4**). After the process is complete, photosynthesis releases oxygen and produces carbohydrate molecules, most commonly glucose. These sugar molecules contain the energy that living things need to survive.

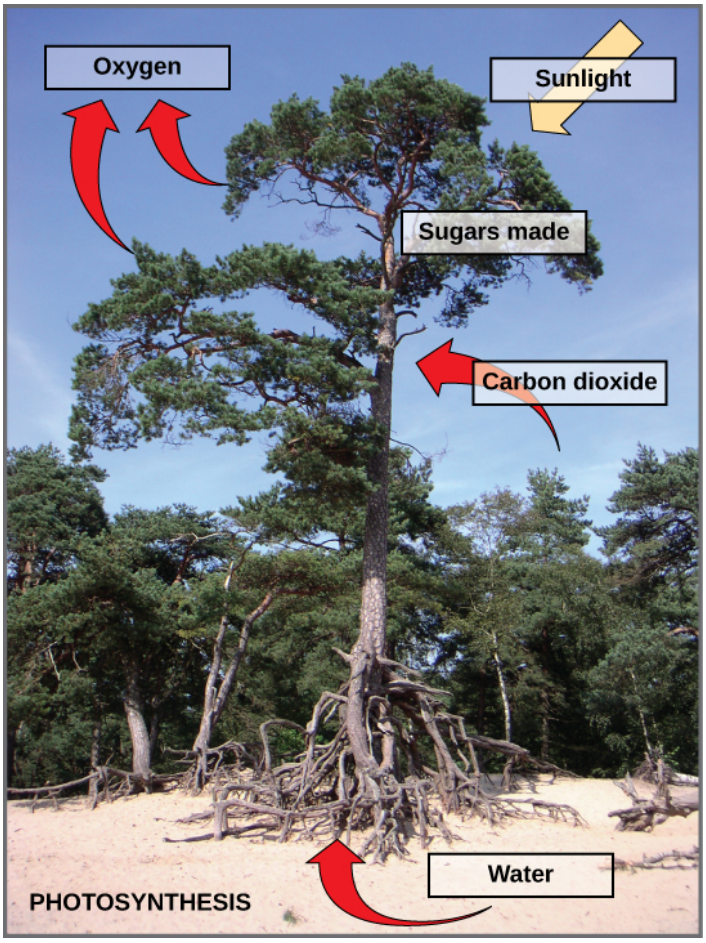


Figure 18.4 Photosynthesis uses solar energy, carbon dioxide, and water to release oxygen and to produce energy-storing sugar molecules.

The complex reactions of photosynthesis can be summarized by the chemical equation shown in **Figure 18.5**.

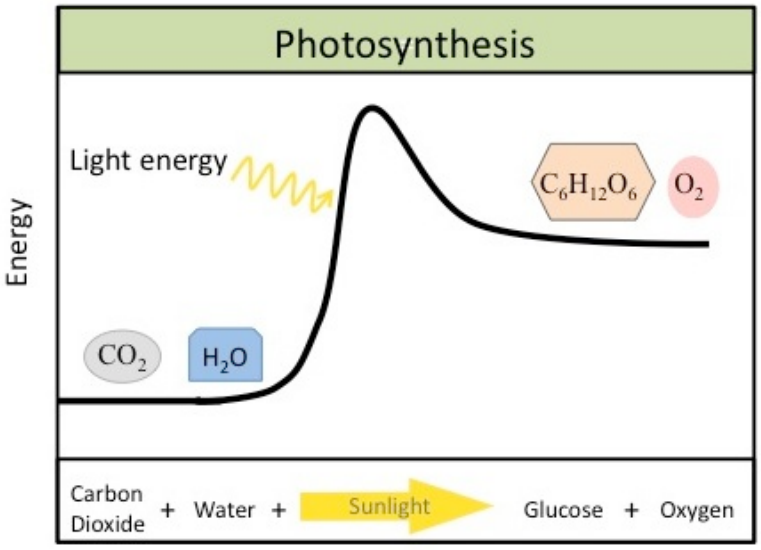


Figure 18.5 The process of photosynthesis can be represented by an equation, wherein carbon dioxide and water produce sugar and oxygen using energy from sunlight. Since the products have a higher energy level than the reactants, photosynthesis is an endergonic reaction. (Image by Robert Bear)

Although the equation looks simple, the many steps that take place during photosynthesis are actually quite complex, as in the way that the reaction summarizing cellular respiration represented many individual reactions. Before learning the details of how photoautotrophs turn sunlight into food, it is important to become familiar with the physical structures involved.

In plants, photosynthesis takes place primarily in leaves, which consist of many layers of cells and have differentiated top and bottom sides. The process of photosynthesis occurs not on the surface layers of the leaf, but rather in a middle layer called the **mesophyll** (Figure 18.6). The gas exchange of carbon dioxide and oxygen occurs through small, regulated openings called **stomata**.

In all autotrophic eukaryotes, photosynthesis takes place inside an organelle called a **chloroplast**. In plants, chloroplast-containing cells exist in the mesophyll. Chloroplasts have a double (inner and outer) membrane. Within the chloroplast is a third membrane that forms stacked, disc-shaped structures called **thylakoids**. Embedded in the thylakoid membrane are molecules of **chlorophyll**, a **pigment** (a molecule that absorbs light) through which the entire process of photosynthesis begins. Chlorophyll is responsible for the green color of plants. The thylakoid membrane encloses an internal space called the thylakoid space. Other types of pigments exist that can carry out photosynthesis, but chlorophyll is by far the most common. As shown in Figure 18.6, a stack of thylakoids is called **granum**, and the space surrounding the granum is called **stroma** (not to be confused with stomata, the openings on the leaves).

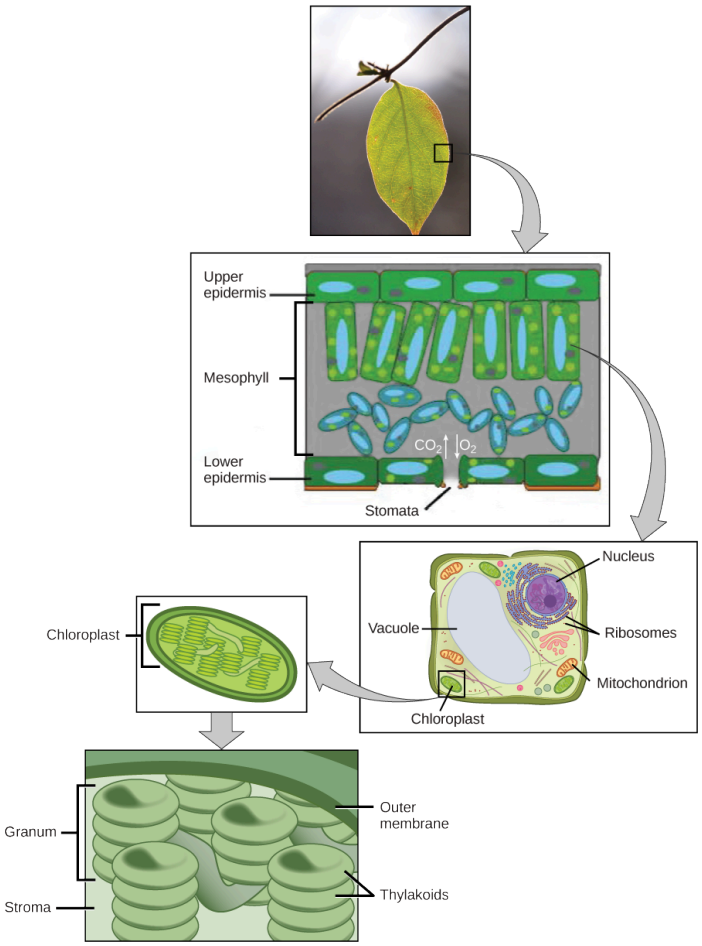


Figure 18.6 Not all cells of a leaf carry out photosynthesis. Cells within the middle layer of a leaf have chloroplasts, which contain the photosynthetic apparatus. (credit "leaf": modification of work by Cory Zanker)

The Two Parts of Photosynthesis

Photosynthesis takes place in two stages: the light-dependent reactions and the Calvin cycle (Figure 18.7). In the **light-dependent reactions**, which take place at the thylakoid membrane, chlorophyll absorbs energy from sunlight and then converts it into chemical energy with the use of water. The light-dependent reactions release oxygen from the hydrolysis of water as a byproduct. In the **Calvin cycle**, which takes place in the stroma, the chemical energy derived from the light-dependent reactions drives both the capture of carbon in carbon dioxide molecules and the subsequent assembly of sugar molecules. The two reactions use carrier molecules to transport the energy from one to the other. The carriers that move energy from the light-dependent reactions to the Calvin cycle reactions can be thought of as “full” because they bring

energy. After the energy is released, the “empty” energy carriers return to the light-dependent reactions to obtain more energy.

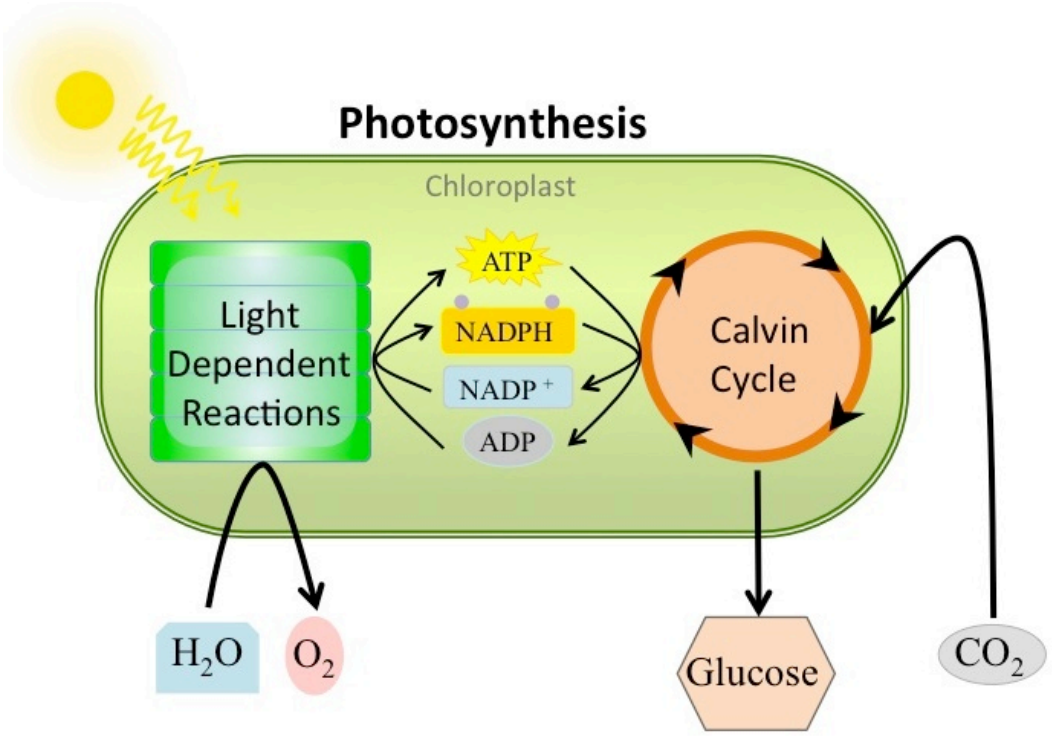


Figure 18.7 The process of photosynthesis is divided to two stages that are linked by the energy and electron carriers ATP and NADPH. (Image by Eva Horne and Robert Bear)

18.2 | The Light-Dependent Reactions

Introduction

> “And God said, Let there be light: and there was light.”

The Book of Genesis 1:3

It is interesting that light is one of the first things mentioned in the Bible, since light indeed has to be present before any of the other creatures, plant or animal, become possible. The light from the sun powers every cell on the planet, allowing plants to make food that the rest of us can also partake of. How can light be used to make food? When a person turns on a lamp, electrical energy becomes light energy. Like all other forms of kinetic energy, light can travel, change form, and be harnessed to do work. In the case of photosynthesis, light energy is converted into chemical energy, which photoautotrophs use to build carbohydrate molecules (**Figure 18.8**). However, autotrophs only use a few specific components of sunlight.

Figure 18.8 Photoautotrophs can capture light energy from the sun, converting it into the chemical energy used to build food molecules. (credit: Gerry Atwell)

What Is Light Energy?

The sun emits an enormous amount of electromagnetic radiation (solar energy). Humans can see only a fraction of this energy, which portion is therefore referred to as "visible light." The manner in which solar energy travels is described as waves. Scientists can determine the amount of energy of a wave by measuring its **wavelength**, the distance between consecutive points of a wave. A single wave is measured from two consecutive points, such as from crest to crest or from trough to trough (**Figure 18.9**).

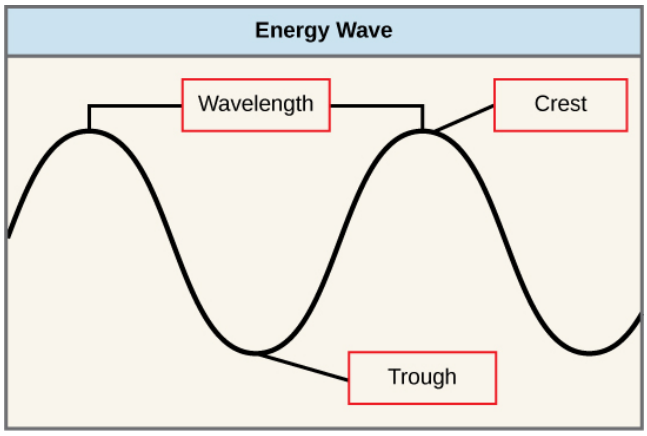


Figure 18.9 The wavelength of a single wave is the distance between two consecutive points of similar position (two crests or two troughs) along the wave.

Visible light constitutes only one of many types of electromagnetic radiation emitted from the sun and other stars. Scientists differentiate the various types of radiant energy from the sun within the electromagnetic spectrum. The **electromagnetic spectrum** is the range of all possible frequencies of radiation (**Figure 18.10**). The difference between wavelengths relates to the amount of energy carried by them.

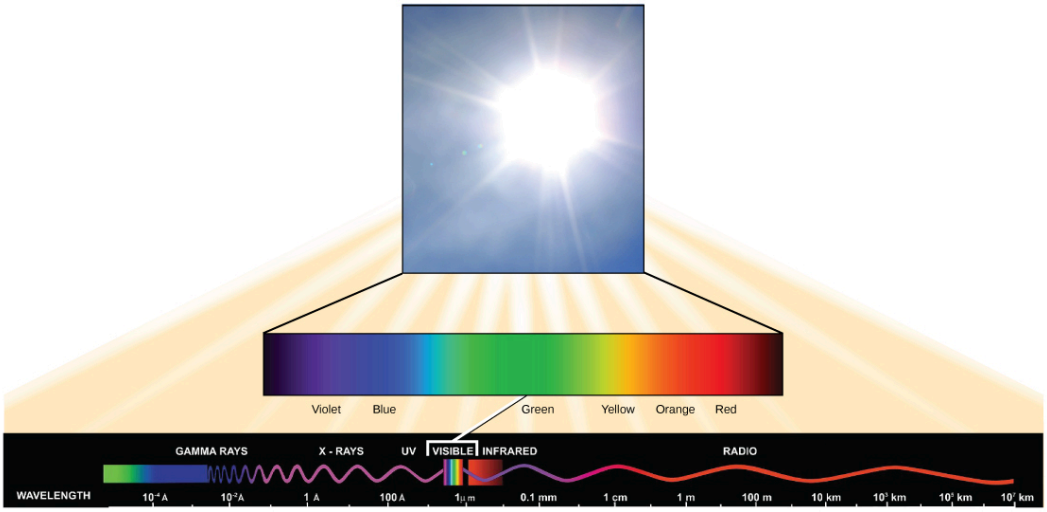


Figure 18.10 The sun emits energy in the form of electromagnetic radiation. This radiation exists at different wavelengths, each of which has its own characteristic energy. All electromagnetic radiation, including visible light, is characterized by its wavelength.

Each type of electromagnetic radiation travels at a particular wavelength. The longer the wavelength (or the more stretched out it appears in the diagram), the less energy is carried. Short, tight waves carry the most energy. This may seem illogical, but think of it in terms of a piece of moving a heavy rope. It takes little effort by a person to move a rope in long, wide waves. To make a rope move in short, tight waves, a person would need to apply significantly more energy.

The electromagnetic spectrum (**Figure 18.10**) shows several types of electromagnetic radiation originating from the sun, including X-rays and ultraviolet (UV) rays. The higher-energy waves can penetrate tissues and damage cells and DNA, explaining why both X-rays and UV rays can be harmful to living organisms.

Absorption of Light

Light energy initiates the process of photosynthesis when pigments absorb the light. Organic pigments, whether in the human retina or the chloroplast thylakoid, have a narrow range of energy levels that they can absorb. Energy levels lower than those represented by red light are insufficient to raise an orbital electron to a populatable, excited (quantum) state. Energy levels higher than those in blue light will physically tear the molecules apart, called bleaching. So retinal pigments can only “see” (absorb) 700 nm to 400 nm light, which is therefore called visible light. For the same reasons, plants pigment molecules absorb only light in the wavelength range of 700 nm to 400 nm; plant physiologists refer to this range for plants as photosynthetically active radiation.

The visible light seen by humans as white light actually exists in a rainbow of colors. Certain objects, such as a prism or a drop of water, disperse white light to reveal the colors to the human eye. The visible light portion of the electromagnetic spectrum shows the rainbow of colors, with violet and blue having shorter wavelengths, and therefore higher energy. At the other end of the spectrum toward red, the wavelengths are longer and have lower energy (**Figure 18.11**).

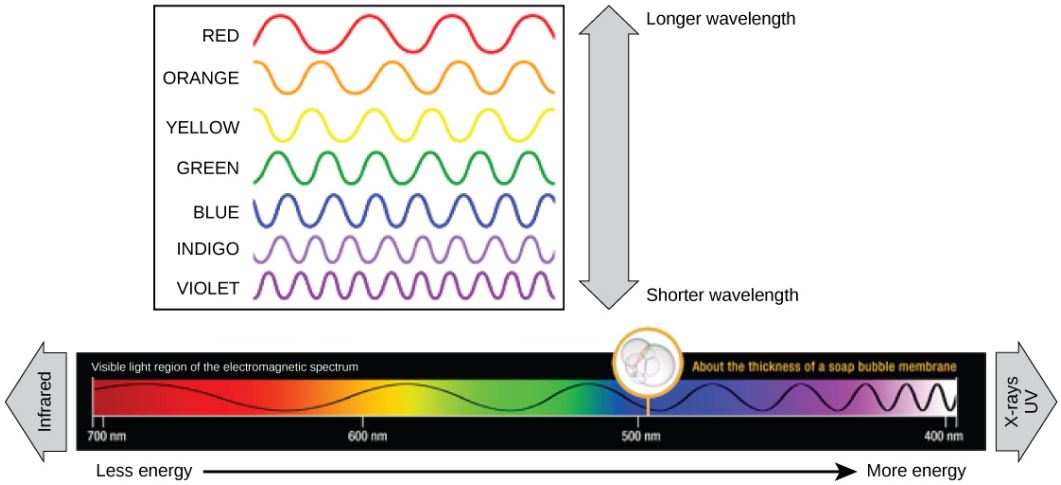


Figure 18.11 The colors of visible light do not carry the same amount of energy. Violet has the shortest wavelength and therefore carries the most energy, whereas red has the longest wavelength and carries the least amount of energy. (credit: modification of work by NASA)

Understanding Pigments

Different kinds of pigments exist, and each has evolved to absorb only certain wavelengths (colors) of visible light. Pigments reflect or transmit the wavelengths they cannot absorb, making them appear in the corresponding color.

Chlorophylls and carotenoids are the two major classes of photosynthetic pigments found in plants and algae; each class has multiple types of pigment molecules. There are five major chlorophylls: *a*, *b*, *c* and *d* and a related molecule found in prokaryotes called bacteriochlorophyll. **Chlorophyll *a*** and **chlorophyll *b*** are found in higher plant chloroplasts and will be the focus of the following discussion.

With dozens of different forms, carotenoids are a much larger group of pigments. The carotenoids found in fruit—such as the red of tomato (lycopene), the yellow of corn seeds (zeaxanthin), or the orange of an orange peel (β-carotene)—are used as advertisements to attract seed dispersers. In photosynthesis, **carotenoids** function as photosynthetic pigments that are very efficient molecules for the disposal of excess energy. When a leaf is exposed to full sun, the light-dependent reactions are required to process an enormous amount of energy; if that energy is not handled properly, it can do significant damage. Therefore, many carotenoids reside in the thylakoid membrane, absorb excess energy, and safely dissipate that energy as heat.

Each type of pigment can be identified by the specific pattern of wavelengths it absorbs from visible light, which is the absorption spectrum. The graph in **Figure 18.12** shows the absorption spectra for chlorophyll *a*, chlorophyll *b*, and a type of carotenoid pigment called β-carotene (which absorbs blue and green light). Notice how each pigment has a distinct set of peaks and troughs, revealing a highly specific pattern of absorption. Chlorophyll *a* absorbs wavelengths from either end of the visible spectrum (blue and red), but not green. Because green is reflected or transmitted, chlorophyll appears green. Carotenoids absorb in the short-wavelength blue region, and reflect the longer yellow, red, and orange wavelengths.

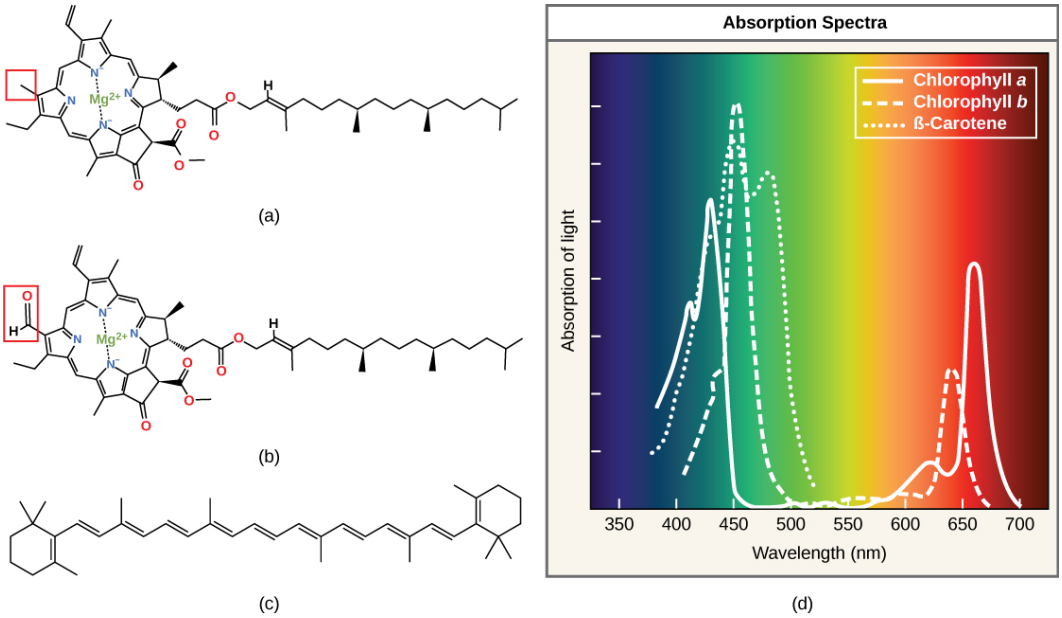


Figure 18.12 (a) Chlorophyll *a*, (b) chlorophyll *b*, and (c) *β*-carotene are hydrophobic organic pigments found in the thylakoid membrane. Chlorophyll *a* and *b*, which are identical except for the part indicated in the red box, are responsible for the green color of leaves. *β*-carotene is responsible for the orange color in carrots. Each pigment has (d) a unique absorbance spectrum.

Many photosynthetic organisms have a mixture of pigments; using them, the organism can absorb energy from a wider range of wavelengths. Not all photosynthetic organisms have full access to sunlight. Some organisms grow underwater where light intensity and quality decrease and change with depth. Other organisms grow in competition for light. Plants on the rainforest floor must be able to absorb any bit of light that comes through, because the taller trees absorb most of the sunlight and scatter the remaining solar radiation (**Figure 18.13**).

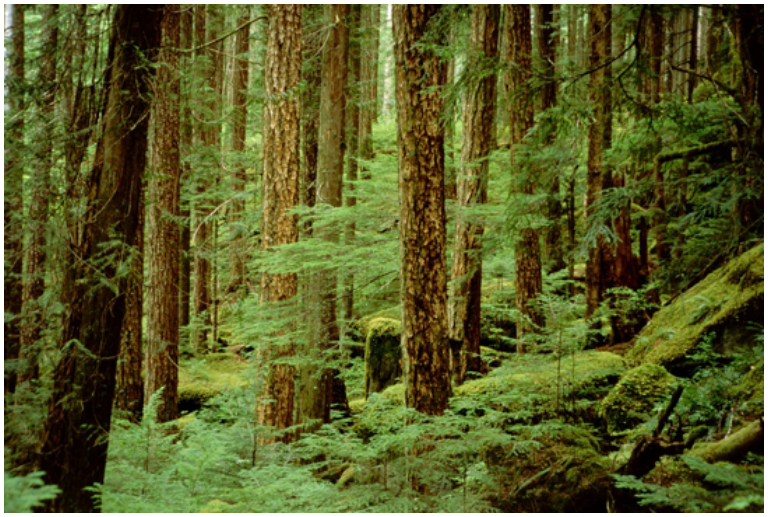


Figure 18.13 Plants that commonly grow in the shade have adapted to low levels of light by changing the relative concentrations of their chlorophyll pigments. (credit: Jason Hollinger)

When studying a photosynthetic organism, scientists can determine the types of pigments present by generating absorption spectra. An instrument called a spectrophotometer can differentiate which wavelengths of light a substance can absorb. Spectrophotometers measure transmitted light and compute from it the absorption. By extracting pigments from leaves and placing these samples into a spectrophotometer, scientists can identify which wavelengths of light an organism can absorb. Additional methods for the identification of plant pigments include various types of chromatography that separate the pigments by their relative affinities to solid and mobile phases.

How Light-Dependent Reactions Work

The overall function of light-dependent reactions is to convert solar energy into chemical energy in the form of NADPH and ATP. This chemical energy supports the Calvin cycle and fuels the assembly of sugar molecules. The light-dependent reactions are depicted in Figure 18.15. Protein complexes and pigment molecules work together to produce NADPH and ATP.

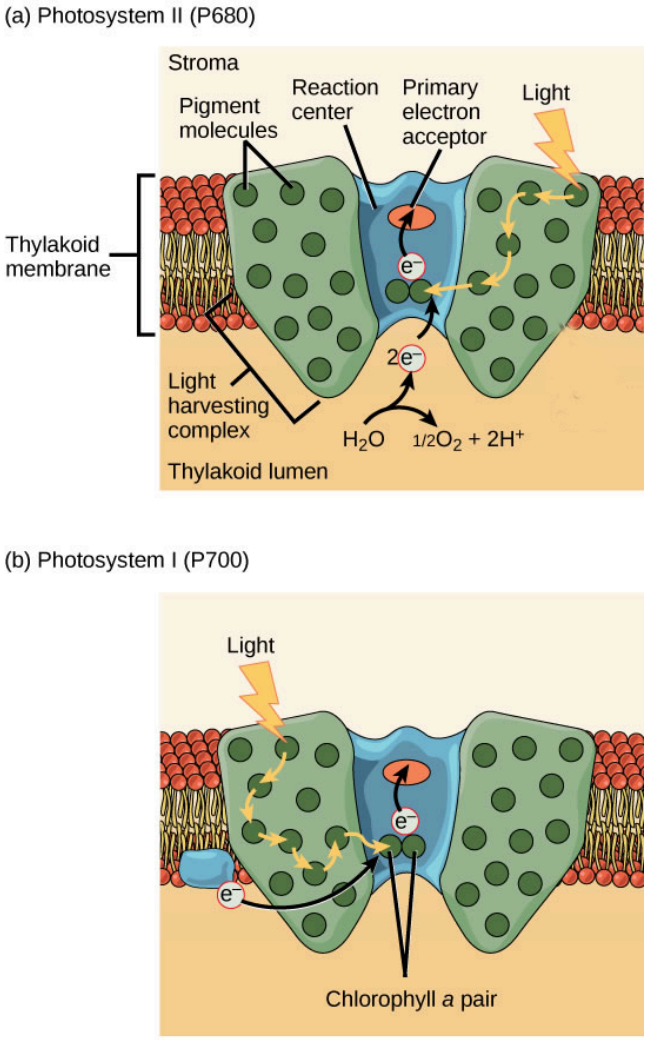


Figure 18.14 A photosystem consists of a light-harvesting complex and a reaction center. Pigments in the light-harvesting complex pass light energy to two special chlorophyll *a* molecules in the reaction center. The light excites an electron from the chlorophyll *a* pair, which passes to the primary electron acceptor. The excited electron must then be replaced. In (a) photosystem II, the electron comes from the splitting of water, which releases oxygen as a waste product. In (b) photosystem I, the electron comes from the chloroplast electron transport chain discussed below.

The actual step that converts light energy into chemical energy takes place in a multiprotein complex called a **photosystem** Figure 18.14, two types of which are found embedded in the thylakoid membrane, **photosystem II** (PSII) and **photosystem I** (PSI) (Figure 18.15). The two complexes differ on the basis of what they oxidize (that is, the source of the low-energy electron supply) and what they reduce (the place to which they deliver their energized electrons).

Both photosystems have the same basic structure; a number of antenna proteins to which the chlorophyll molecules are bound surround the **reaction center** where the photochemistry takes place. Each photosystem is serviced by the light-harvesting complex, which passes energy from sunlight to the reaction center; it consists of multiple antenna proteins that contain a mixture of 300–400 chlorophyll *a* and *b* molecules as well as other pigments like carotenoids. The absorption of a single **photon** or distinct quantity or "packet" of light by any of the chlorophylls pushes that molecule into an excited state. In short, the light energy has now been captured by biological molecules but is not stored in any useful form yet. The energy is transferred from chlorophyll to chlorophyll until eventually (after about a millionth of a second), it is delivered to the reaction center. Up to this point, only energy has been transferred between molecules, not electrons.

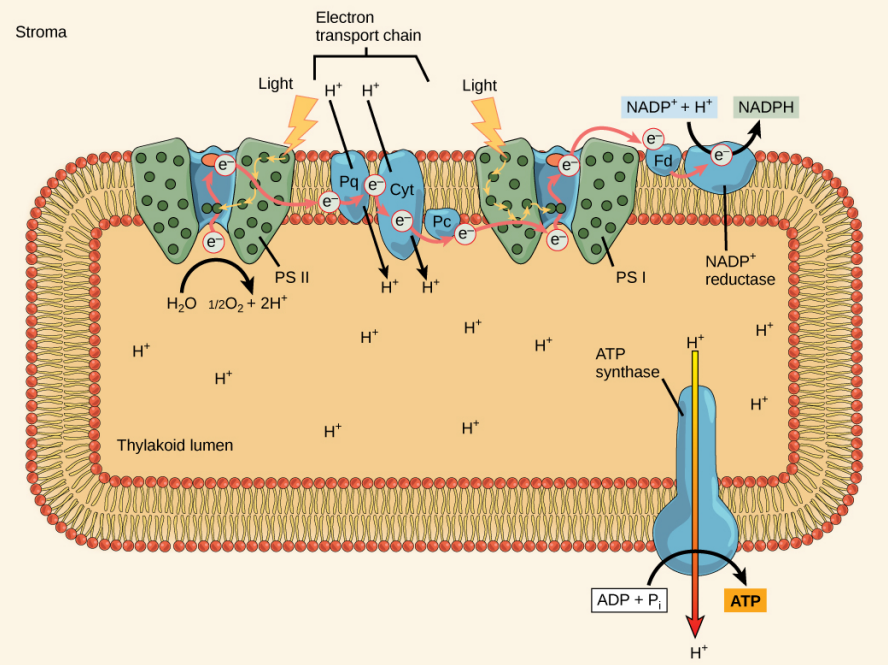


Figure 18.15 In the photosystem II (PSII) reaction center, energy from sunlight is used to extract electrons from water. The electrons travel through the chloroplast electron transport chain to photosystem I (PSI), which reduces $NADP^+$ to NADPH. The electron transport chain moves protons across the thylakoid membrane into the lumen. At the same time, splitting of water adds protons to the lumen, and reduction of NADPH removes protons from the stroma. The net result is a low pH in the thylakoid lumen, and a high pH in the stroma. ATP synthase uses this electrochemical gradient to make ATP.

The reaction center contains a pair of chlorophyll *a* molecules with a special property. Those two chlorophylls can undergo oxidation upon excitation; they can actually give up an electron in a process called a photoact. It is at this step in the reaction center, this step in photosynthesis, that light energy is converted into an excited electron. All of the subsequent steps involve getting that electron onto the electron carrier $NADP^+$ for delivery to the Calvin cycle where the electron is deposited onto carbon for long-term storage in the form of a carbohydrate. PSII and PSI are two major components of the photosynthetic **electron transport chain**, which also includes the cytochrome complex, a group of reversibly oxidizable and reducible proteins that forms part of the electron transport chain between PSII and PSI.

The reaction center of PSII (called **P680**) delivers its high-energy electrons, one at a time, to a series of proteins and electron carriers (or primary electron acceptors, which are pigments or other organic molecules in the reaction center that accept energized electrons from the reaction center) that sits between it and PSI. P680's missing electron is replaced by extracting a low-energy electron from water; thus, water is split and PSII is re-reduced after every photoact. Splitting one H_2O molecule releases two electrons, two hydrogen atoms, and one atom of oxygen. Splitting two molecules is required to form one molecule of diatomic O_2 gas. About 10 percent of the oxygen is used by mitochondria in the leaf to support oxidative phosphorylation. The remainder escapes to the atmosphere where it is used by aerobic organisms to support respiration.

As electrons move through the proteins that reside between PSII and PSI, they lose energy. That energy is used to move hydrogen atoms from the stromal side of the membrane to the thylakoid lumen. Those hydrogen atoms, plus the ones produced by splitting water, accumulate in the thylakoid lumen and will be used synthesize ATP in a later step. Because the electrons have lost energy prior to their arrival at PSI, they must be re-energized by PSI, hence, another photon is absorbed by the PSI antenna. That energy is relayed to the PSI reaction center (called **P700**). P700 is oxidized and sends a high-energy electron to $NADP^+$ to form NADPH. Thus, PSII captures the energy to make ATP, and PSI captures the energy to reduce $NADP^+$ into NADPH. The two photosystems work in concert, in part, to guarantee that the production of NADPH will roughly equal the production of ATP. Other mechanisms exist to fine tune that ratio to exactly match the chloroplast's constantly changing energy needs.

Generating an Energy Carrier: ATP

Chemiosmosis

In chemiosmosis, the free energy from the series of redox reactions just described is used to pump hydrogen ions (protons) across the membrane. The uneven distribution of H^+ ions across the thylakoid membrane establishes both concentration and electrical gradients (thus, an electrochemical gradient), owing to the hydrogen ions' positive charge and their aggregation on one side of the membrane.

If the thylakoid membrane were open to diffusion by the hydrogen ions, the ions would tend to diffuse back across into the matrix, driven by their electrochemical gradient. Recall that many ions cannot diffuse through the nonpolar regions of phospholipid membranes without the aid of ion channels. Similarly, hydrogen ions in the thylakoid lumen can only pass through the thylakoid membrane through an integral membrane protein called ATP synthase (**Figure 18.16**). This complex protein acts as a tiny generator, turned by the force of the hydrogen ions diffusing through it, down their electrochemical gradient. The turning parts of this molecular machine facilitates the addition of a phosphate to ADP, forming ATP, using the potential energy of the hydrogen ion gradient.

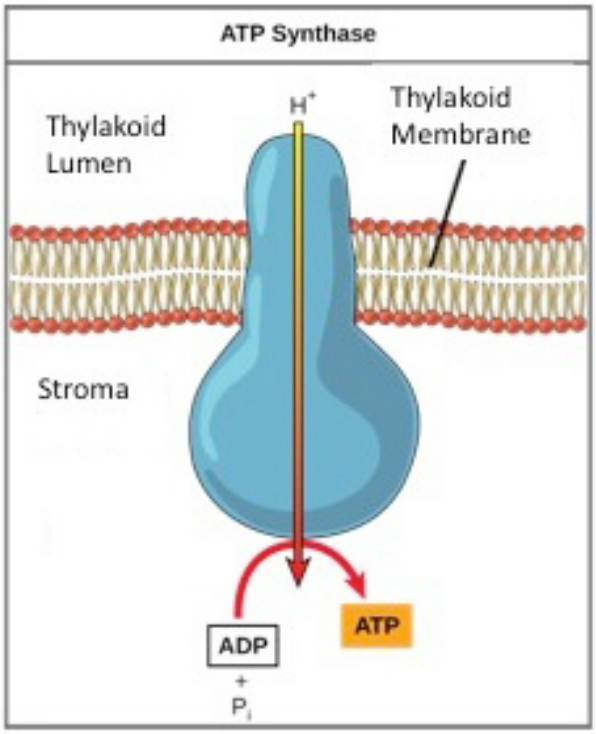


Figure 18.16 ATP synthase is a complex, molecular machine that uses a proton (H^+) gradient to form ATP from ADP and inorganic phosphate (Pi). (Credit: modification of work by Klaus Hoffmeier)

As a review, the buildup of hydrogen ions inside the thylakoid lumen creates a concentration gradient. The passive diffusion of hydrogen ions from high concentration (in the thylakoid lumen) to low concentration (in the stroma) is harnessed to create ATP. The ions build up energy because of diffusion and because they all have the same electrical charge, repelling each other. To release this energy, hydrogen ions will rush through any opening, similar to water jetting through a hole in a dam. In the thylakoid, that opening is a passage through a specialized protein channel called the ATP synthase. The energy released by the hydrogen ion stream allows ATP synthase to attach a third phosphate group to ADP, which forms a molecule of ATP (**Figure 18.16**). The flow of hydrogen ions through ATP synthase is called chemiosmosis because the ions move from an area of high to an area of low concentration through a semi-permeable structure.

Overview of Light-Dependent Reaction

Light energy is harvested and transformed into short term chemical energy during the light-dependent reactions. The pigments in the reaction center of photosytem II absorb light energy and excite electrons into the electron transport chain.

To replace the electrons excited in photosystem II, water is spilt releasing electrons, H^+ ions and oxygen gas. As the electrons move through the electron transport chain, energy, release by the movement of electrons, is used to produce a H^+ ion gradient inside the thylakoid. The process of chemiosmosis uses the H^+ concentration to produce ATP. After flowing through the electron transport chain, the electrons enter photosystem I. The reaction center of photosystem I absorbs more light energy and excites electrons. These energized electrons reduce $NADP^{+,}$ to NADPH. See (**Figure 18.17**)below for a visual representation of the light-dependent reactions.

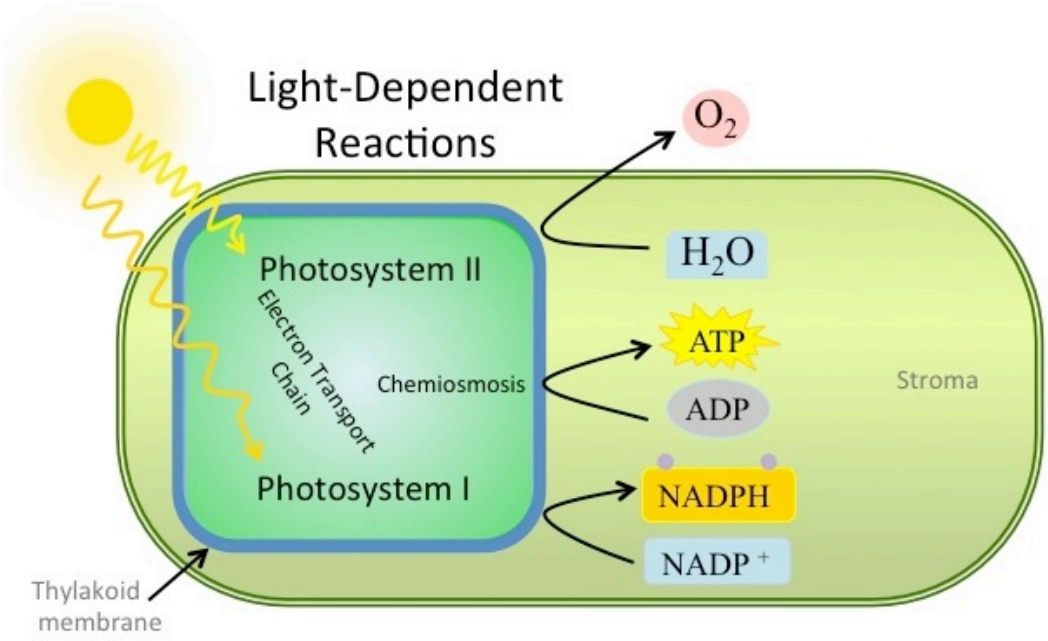


Figure 18.17 This is an image of the Light-Dependent Reactions of photosynthesis. (Image by Eva Horne and Robert Bear)

18.3 | Calvin Cycle

> "By blending water and minerals from below with sunlight and CO_2 from above, green plants link the earth to the sky. We tend to believe that plants grow out of the soil, but in fact most of their substance comes from the air."

Fritjof Capra, *The Web of Life: A New Scientific Understanding of Living Systems*, (1997)

After the energy from the sun is converted into chemical energy and temporarily stored in ATP and NADPH molecules, the cell has the fuel needed to build carbohydrate molecules for long-term energy storage. The products of the light-dependent reactions, ATP and NADPH, have lifespans in the range of millionths of seconds, whereas the products of the light-independent reactions (carbohydrates and other forms of reduced carbon) can survive for hundreds of millions of years. Carbohydrates, like the cellulose that makes up the bulk of most plants, obviously contain carbon. Where does the carbon come from? It comes from the air, in the form of carbon dioxide, the gas that is a waste product of respiration in microbes, fungi, plants, and animals.

The Calvin Cycle

In plants, carbon dioxide (CO_2) enters the leaves through stomata, where it diffuses over short distances through intercellular spaces until it reaches the mesophyll cells. Once in the mesophyll cells, CO_2 diffuses into the stroma of

the chloroplast—the site of light-independent reactions of photosynthesis. These reactions actually have several names associated with them. Another term, the **Calvin cycle**, is named for the man who discovered it, and because these reactions function as a cycle. Others call it the Calvin-Benson cycle to include the name of another scientist involved in its discovery. The most outdated name is dark reactions (the term originally used by Melvin Calvin, who got the Nobel Prize for elucidating these reactions), because light is not directly required (**Figure 18.18**). However, the term dark reactions can be misleading because it implies incorrectly that the reaction only occurs at night or is independent of light, which is why it has faded from everyday usage.

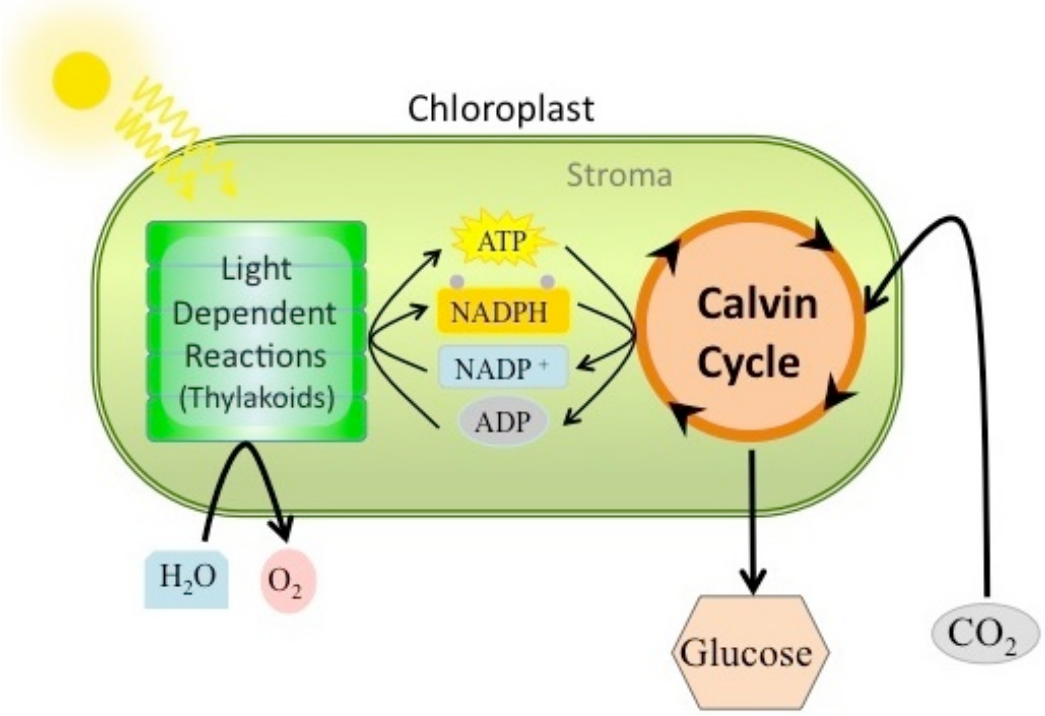


Figure 18.18 Light reactions harness energy from the sun to produce chemical bonds, ATP, and NADPH. These energy-carrying molecules are made in the stroma where carbon fixation takes place. Work by Eva Horne and Robert A. Bear

The light-independent reactions of the Calvin cycle can be organized into three basic stages: fixation, reduction, and regeneration.

Stage 1: Fixation

In the stroma, in addition to CO_2, two other components are present to initiate the light-independent reactions: an enzyme called ribulose bisphosphate carboxylase (RuBisCO), and three molecules of ribulose bisphosphate (RuBP), as shown in **Figure 18.19**. RuBP has five atoms of carbon, flanked by two phosphates.

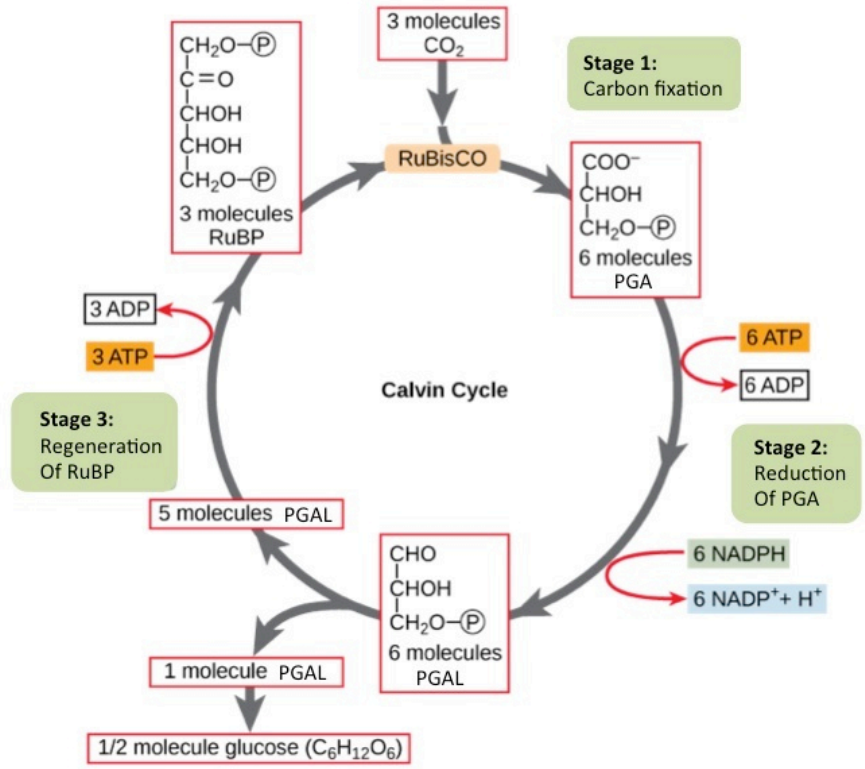


Figure 18.19 The Calvin cycle has three stages. In stage 1, the enzyme RuBisCO incorporates carbon dioxide into an organic molecule, PGA. In stage 2, the organic molecule is reduced using electrons supplied by NADPH. In stage 3, RuBP, the molecule that starts the cycle, is regenerated so that the cycle can continue. Only one carbon dioxide molecule is incorporated at a time, so the cycle must be completed three times to produce a single three-carbon PGAL molecule, and six times to produce a six-carbon glucose molecule. (Original by OpenStax Modified by Robert A. Bear)

RuBisCO catalyzes a reaction between CO_2 and RuBP. For each CO_2 molecule that reacts with one RuBP, two molecules of phosphoglycerate (PGA) form. PGA has three carbons and one phosphate. Each turn of the cycle involves only one RuBP and one carbon dioxide and forms two molecules of PGA. The number of carbon atoms remains the same, as the atoms move to form new bonds during the reactions (3 atoms from $3CO_2$ + 15 atoms from 3RuBP = 18 atoms in 3 atoms of PGA). This process is called **carbon fixation**, because CO_2 is "fixed" from an inorganic form into organic molecules.

Stage 2: Reduction

ATP and NADPH are used to convert the six molecules of PGA into six molecules of a chemical called phosphoglyceraldehyde (PGAL). That is a reduction reaction because it involves the gain of electrons (from NADPH) by PGA. Recall that a **reduction** is the gain of an electron by an atom or molecule. Six molecules of both ATP and NADPH are used; making glucose is obviously an energy-intensive activity. For ATP, energy is released with the loss of the terminal phosphate atom, converting it into ADP; for NADPH, both energy and a hydrogen atom are lost, converting it into $NADP^+$. Both of these molecules return to the nearby site of the light-dependent reactions to be reused and re-energized.

Stage 3: Regeneration

Interestingly, at this point, only one of the PGAL molecules leaves the Calvin cycle and is sent to the cytoplasm to contribute to the formation of other compounds needed by the plant. Because the PGAL exported from the chloroplast has three carbon atoms, it takes three "turns" of the Calvin cycle to fix enough net carbon to export one PGAL. But each turn makes two PGAL, thus three turns make six PGAL. One is exported while the remaining five PGAL molecules remain in the cycle and are used to regenerate RuBP, which enables the system to prepare for more CO_2 to be fixed. Three more molecules of ATP are used in these regeneration reactions.

Overview of Calvin Cycle

During the Calvin cycle, energy and electrons harvested in the light-dependent reactions are used to produce carbohydrates i.e. glucose. There are three stages in the Calvin cycle. The first stage is carbon fixation, CO_2 from the atmosphere is attached to an organic molecule RuBP during this stage. The second stage is the carbon reduction, the energy and electrons in ATP and NADPH are used to produce carbohydrates (glucose). The last stage is the regeneration stage, energy from ATP

is used to regenerate the first substrate of the cycle (RUBP). See (Figure 18.20) below for a review of the Calvin Cycle.

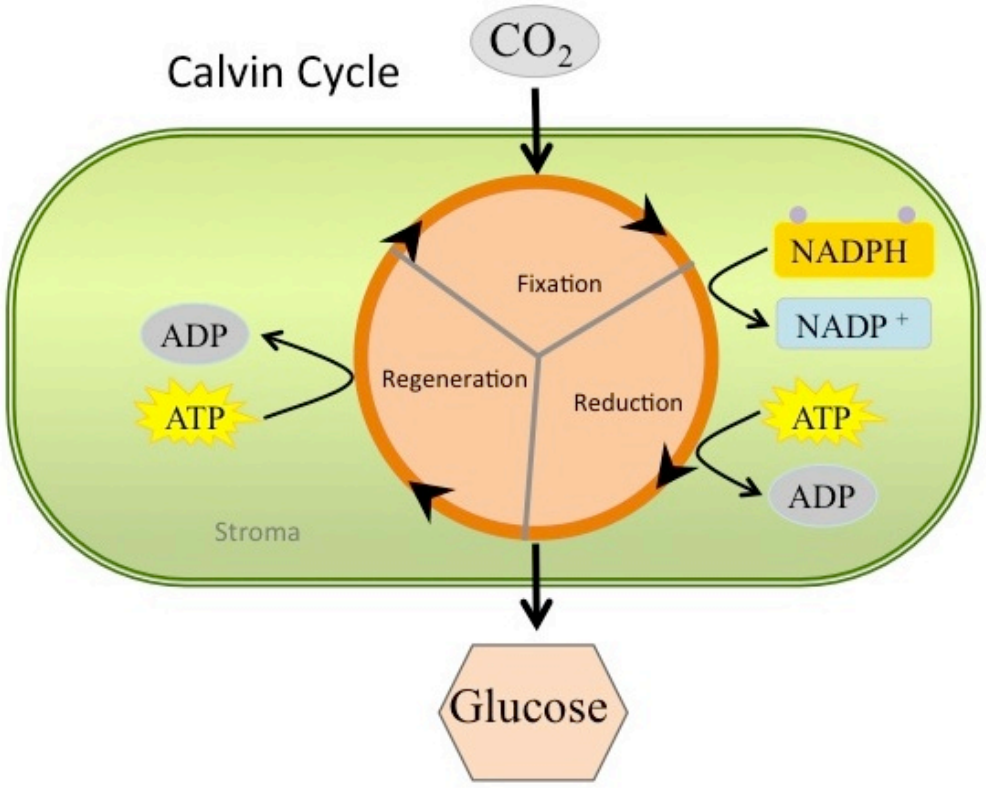


Figure 18.20 This image represents the Calvin Cycle. (Image by Eva Horne and Robert Bear)

evolution CONNECTION

Photosynthesis

During the evolution of photosynthesis, a major shift occurred from the bacterial type of photosynthesis that involves only one photosystem and is typically anoxygenic (does not generate oxygen) into modern oxygenic (does generate oxygen) photosynthesis, employing two photosystems. This modern oxygenic photosynthesis is used by many organisms—from giant tropical leaves in the rainforest to tiny cyanobacterial cells—and the process and components of this photosynthesis remain largely the same. Photosystems absorb light and use electron transport chains to convert energy into the chemical energy of ATP and NADH. The subsequent light-independent reactions then assemble carbohydrate molecules with this energy.

Photosynthesis in desert plants has evolved adaptations that conserve water. In the harsh dry heat, every drop of water must be used to survive. Because stomata must open to allow for the uptake of CO_2, water escapes from the leaf during active photosynthesis. Desert plants have evolved processes to conserve water and deal with harsh conditions. A more efficient use of CO_2 allows plants to adapt to living with less water. Some plants such as cacti (Figure 18.21) can prepare materials for photosynthesis during the night by a temporary carbon fixation/storage process, because opening the stomata at this time conserves water due to cooler temperatures. In addition, cacti have evolved the ability to carry out low levels of photosynthesis without opening stomata at all, an extreme mechanism to face extremely dry periods.

Figure 18.21 The harsh conditions of the desert have led plants like these cacti to evolve variations of the light-independent reactions of photosynthesis. These variations increase the efficiency of water usage, helping to conserve water and energy. (credit: David A. Rintoul)

Overview of Photosynthesis

Photosynthesis converts light energy to chemical energy in two stages, the light-dependent reactions and the Calvin cycle. By exploring these two sets of reactions, we learned how photons of light energy are turned into food by photosynthesis. The light-dependent reactions harvest the light energy to make ATP and to transfer electrons from H_2O to $NADP^+$ forming NADPH and Oxygen gas. The energy and electrons in ATP and NADPH are used in the Calvin cycle to produce glucose from carbon dioxide. The sunlight energy entering the chloroplasts becomes stored as the chemical bonds in the organic molecules. See (**Figure 18.22**) below for a review of photosynthesis.

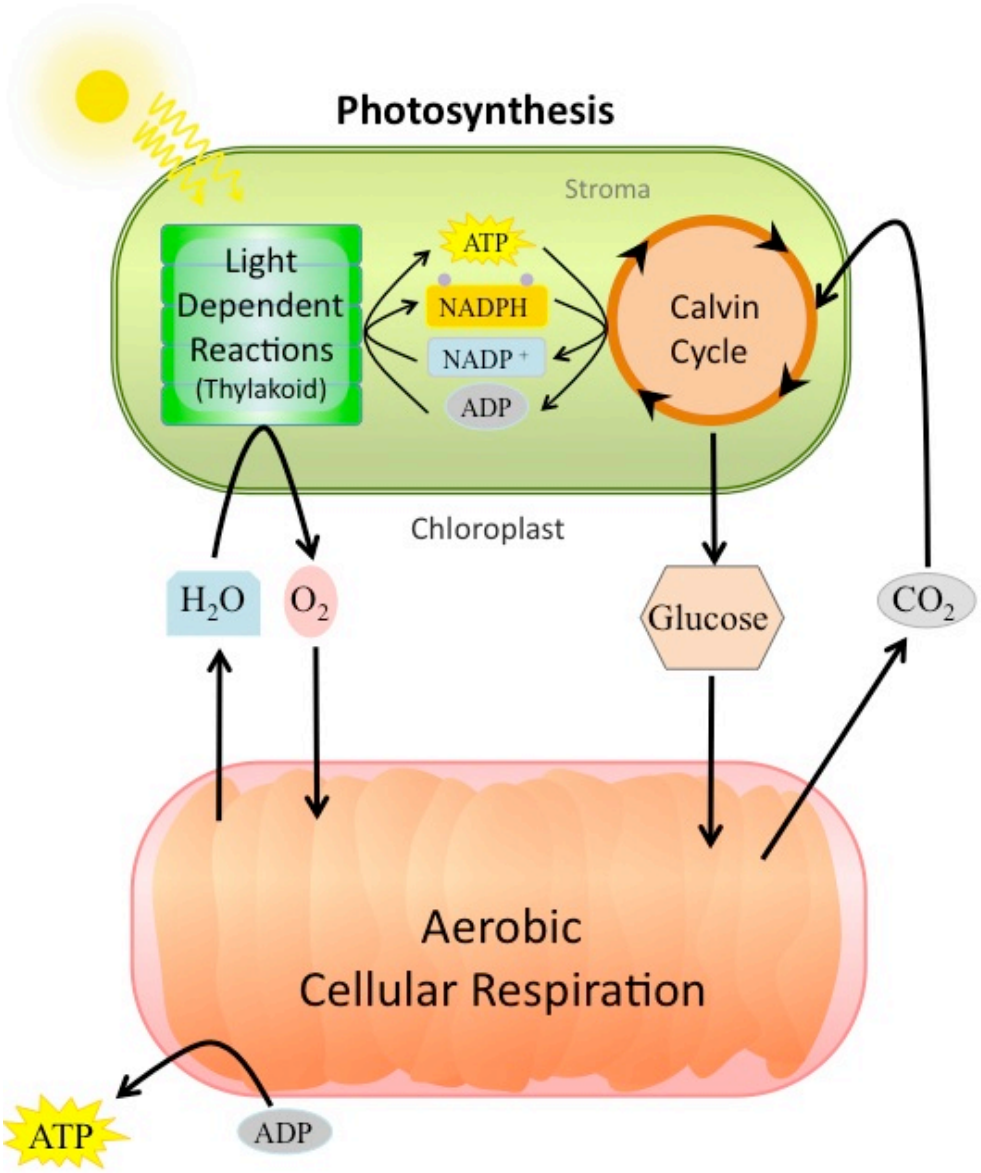


Figure 18.22 The process of photosynthesis is divided to two stages that are linked by the energy and electron carriers ATP and NADPH. The light-dependent reactions split water and releases oxygen as a byproduct, and these reactions convert light energy to chemical energy (ATP and NADPH). The Calvin cycle uses the energy in ATP and NADPH and produces carbohydrates by fixing CO_2 a byproduct of aerobic cellular respiration. (Image by Eva Horne and Robert Bear)

What is the fate of the carbohydrates produced by photosynthesis? About 50% of the carbohydrates are used by the plant for aerobic cellular respiration in their mitochondria. The other 50% of the carbohydrates are the building blocks for the biological macromolecules the make up plant cells that you learned about in module 3. As you may realize, these biological macromolecules are the food we eat, and you are what you eat. Taking this a bit further, all the activities you do from reading this text to sleeping require energy and that energy comes from the Sun.

19 | CELLULAR RESPIRATION

19.1 | Overview of Cellular Respiration

This is a general overview of cellular respiration.

Introduction

> "Surely the mitochondrion that first entered another cell was not thinking about the future benefits of cooperation and integration; it was merely trying to make its own living in a tough Darwinian world."

Stephen Jay Gould, in *Wonderful Life: the Burgess Shale and the Nature of History*, (1990)

All living organisms require energy, and for all organisms this energy comes from the chemical energy found in compounds that they acquire from their environment. The mitochondrion, a descendent of an aerobically-respiring bacteria, is the site of energy generation in eukaryotes. As we learned previously, the process of photosynthesis uses solar energy (sunlight) and converts this energy into chemical energy in the form of carbohydrates. In order for the chemical energy in the carbohydrates to be made available to do cellular work, the energy must be converted into a useable form known as ATP. Adenosine Triphosphate is the energy currency of the cell, and everything you do from walking down the street to reading this book requires energy in the form of ATP. Organisms need a constant supply of ATP, and the potential energy stored in food is the source of energy to meet this need. By connecting all this together, you should realize that your daily activities are fueled by the energy from the sun and that even on the cellular level nutrients cycle and energy flows (**Figure 19.1**).

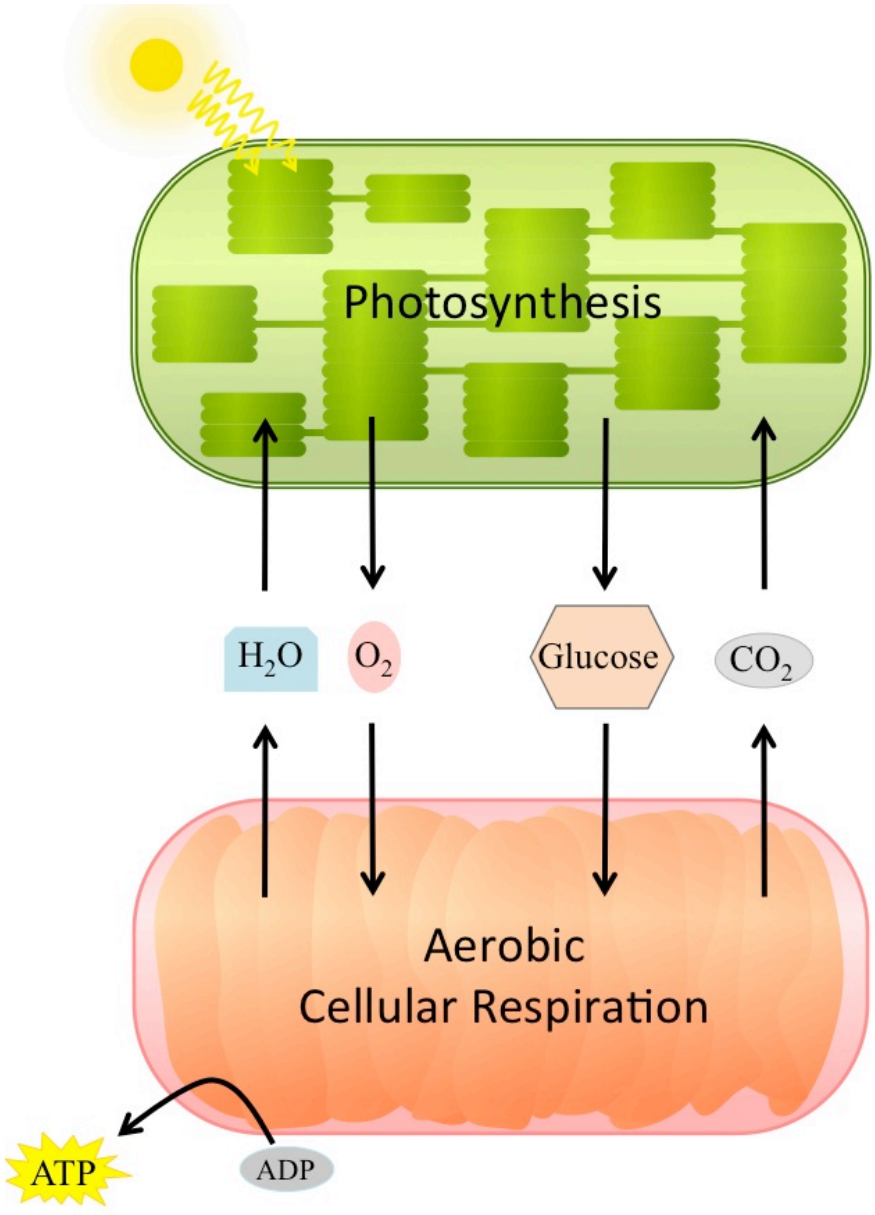


Figure 19.1 This image illustrates the relationship between photosynthesis and cellular respiration. (Image by Eva Horne and Robert Bear)

All organisms need ATP, but not all organisms use the same pathways to generate ATP from the food that is consumed. **Aerobic cellular respiration**, the main subject of this chapter, uses oxygen (O_2) and glucose to generate ATP. Organisms (plants, animals, fungi and microbes) that live in an oxygen (O_2) rich environment use this process to generate ATP. The overall equation for aerobic cellular respiration is the reverse of photosynthesis, is an exergonic reaction, and supplies the ATP for cellular functions (**Figure 19.2**).

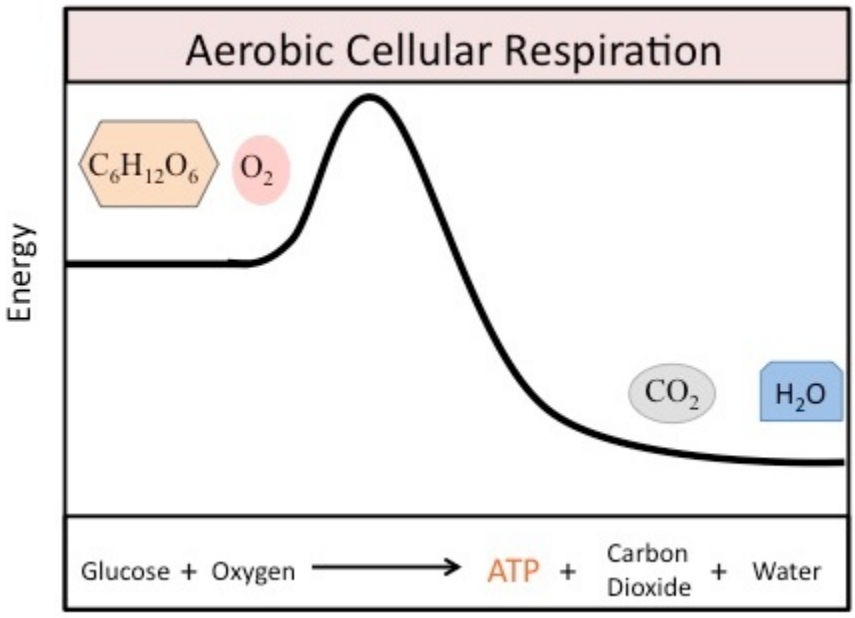


Figure 19.2 This image illustrates the overall equation for aerobic cellular respiration and how the amounts of free energy differs between the reactants and the products. (Image by Robert Bear)

As the aerobic cellular respiration equation shows (**Figure 19.2**), an organism needs to acquire the O_2 from its surroundings and to get rid of the CO_2 that is produced. The acquisition of O_2 and the release of CO_2 is accomplished in a variety of ways. In single celled organisms, the movement of O_2 and CO_2 (gas exchange) is done by simple diffusion. However, in complex organisms there are specialized organs that allow for gas exchange; for example, gills in aquatic organisms and lungs in terrestrial animals.

A common misconception is that plants do not undergo cellular respiration because they make their own energy by photosynthesis. Plants do perform cellular respiration using the carbohydrates produced via photosynthesis; this occurs in tissues that are not photosynthetically active (e.g., roots), as well as in leaves and stems. Approximately half of the glucose produced by photosynthesis is consumed by the plant, mostly to generate ATP during aerobic cellular respiration. Other uses of glucose in the plant include synthesis of cell walls, starch, and other plant carbohydrates. So, plants harvest light energy via photosynthesis, making carbohydrates, and then they use the energy stored in those carbohydrates to perform various cellular functions. This is the reason why they are called **autotrophs**, or self feeders.

Some single-celled organisms use **anaerobic metabolism** to extract energy from biological molecules; this process occurs in the absence of oxygen. In this chapter, we will explore one type of anaerobic metabolism called fermentation. You may already be familiar with a one type of fermentation, lactic acid fermentation, especially if you have recently over-exerted your muscles. Anaerobic metabolism is used by many organisms to produce ATP when oxygen is not available and thus pathways which require oxygen cannot be used. The amount of ATP produced by fermentation is much less then that produced by aerobic cellular respiration, so there is a cost and benefit associated with organisms utilizing fermentation.

Summary of Aerobic Cellular Respiration

Aerobic cellular respiration (**Figure 19.3**) is series of linked chemical reactions that can be best understood if it is separated into four stages. These are **glycolysis**, pyruvate oxidation, the **Krebs Cycle**, and **oxidative phosphorylation**. Similar to photosynthesis, cellular respiration uses a series of oxidation-reduction reactions. During these reactions, electrons are stripped from the chemical bonds of the original glucose molecule and eventually added to oxygen, via a series of intermediate steps. This series of reactions releases small amounts of energy at each step; this energy is used to drive the formation of ATP. This section is a brief introduction to the stages of aerobic respiration with more detail to follow in the chapter.

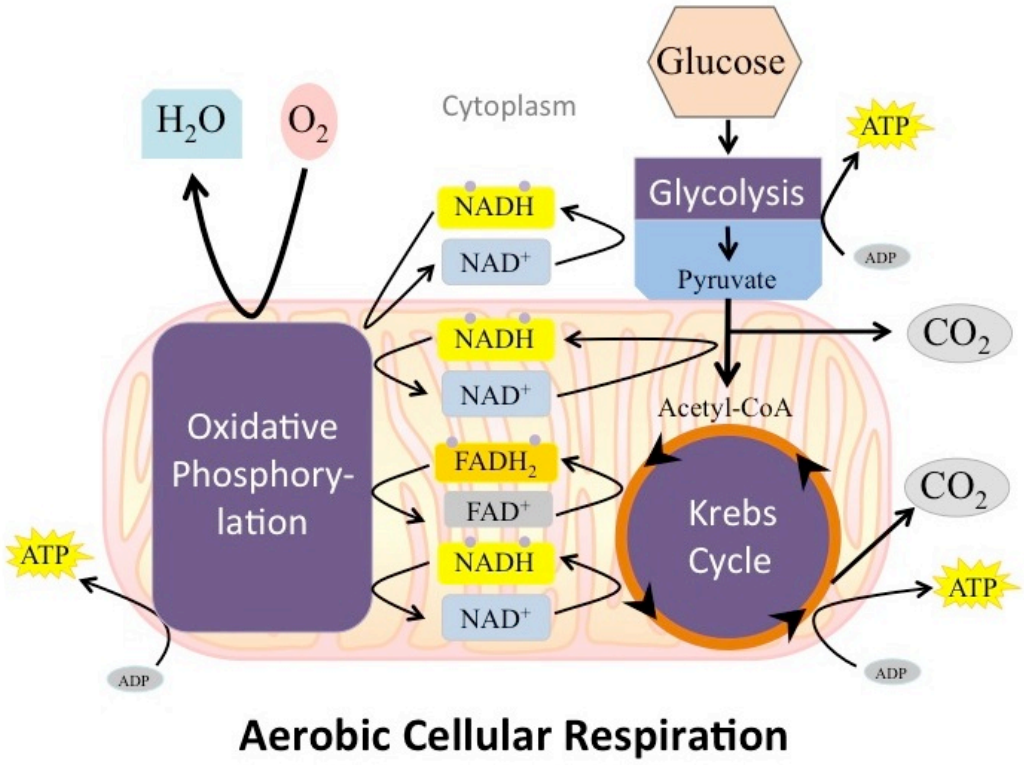


Figure 19.3 This image illustrates aerobic cellular respiration. (Image by Eva Horne and Robert Bear)

The first stage of cellular respiration is called **Glycolysis** and occurs in the cytoplasm of the cell. During glycolysis, 1 glucose molecule (with 6 carbon atoms) is broken down into 2 pyruvate molecules (with three carbon atoms each). This is accompanied by the production of a few ATP molecules and the storage of some high-energy electrons on the electron carrier NADH. Note that no O_2 is needed for this set of reactions, which means that glycolysis can proceed in the absence of oxygen.

The second stage is a short series of reactions called the **oxidation of pyruvate** during which pyruvate (3 carbon atoms) is converted to acetyl-CoA (two carbon atoms), accompanied by the production of CO_2 (one carbon atom). This process occurs on the mitochondrial inner membrane, and as a result the acetyl-CoA is formed inside the mitochondria. Pyruvate is made in the cytoplasm, and this step moves the next compound in the pathway into the mitochodria. This is critical, since all subsequent steps in the pathway occur within the mitochondria. The other important event of this stage is the addition of high-energy electrons to NAD^+, generating another molecule of the electron carrier NADH.

Acetyl-CoA enters into the **Krebs cycle**, a series of mitochondrial reactions that completes the breakdown of the original glucose, thereby releasing CO_2. In this third stage of the process, energy is harvested in the form of high-energy electrons being used to generate NADH as well as another high-energy electron carrier, $FADH_2$. The reactions of the Krebs cycle also produce a small amount of ATP.

So far, a minimal amount of ATP has been produced, but a lot of energy has been stored in the electron carriers NADH and $FADH_2$. In the final stage of aerobic cellular respiration, **Oxidative Phosphorylation**, a series of enzymes known as the **electron transport chain** uses those high-energy electrons to produce a large amount of ATP. The high energy electrons harvested in the first three stages, and ferried by electron carriers (NADH and $FADH_2$) to the electron transport chain, are used to produce large amounts of ATP vir the mitochondrial membrane protein known as the **ATP synthase**. During this final stage is also when atmospheric oxygen is used as the final electron and hydrogen ion acceptor, in a reaction which produces water. The need for O_2 in this final step means that these reactions are part of aerobic cellular respiration.

Location and Structures of Aerobic Cellular Respiration

All eukaryotic cells (protists, fungi, plants and animals) have mitochondria, and mitochondria are often called the power plants of the cell because these organelles produce a large amount of ATP. As you may remember from a previous module, the mitochondrion is an organelle that is hypothesized to have originated as an endosymbiotic aerobic bacteria. Some of the evidence for this hypothesis comes from the relationship of the functional parts of the mitochondria (**Figure 19.4**) to

the structure of a typical aerobic bacteria. There is an **outer membrane** which defines the organelle and represents the membrane which enveloped the bacteria when it was taken into the cell via endocytosis. The inner membrane represents the plasma membrane of the bacteria; the inner and outer membranes together form the **intermembrane space**. The **inner membrane** is highly folded; these folds are called **cristae**. The extensive folding increases the surface area for the numerous electron transport chain enzymes and the ATP synthases that are used to make ATP. In bacteria all of these enzymes are packed into the plasma membrane, as one would expect if the endosymbiotic hypothesis is correct. The production of ATP is driven by a concentration gradient between the outer and inner compartment; in aerobic bacteria this concentration gradient is between the inside and the outside of the cell. The innermost compartment, derived from the cytoplasm of the ancestral bacteria, is called the **matrix**, and this compartment (just like the cytoplasm of today's bacteria) contains ribosomes and DNA; It is also the location of the Krebs Cycle reactions.

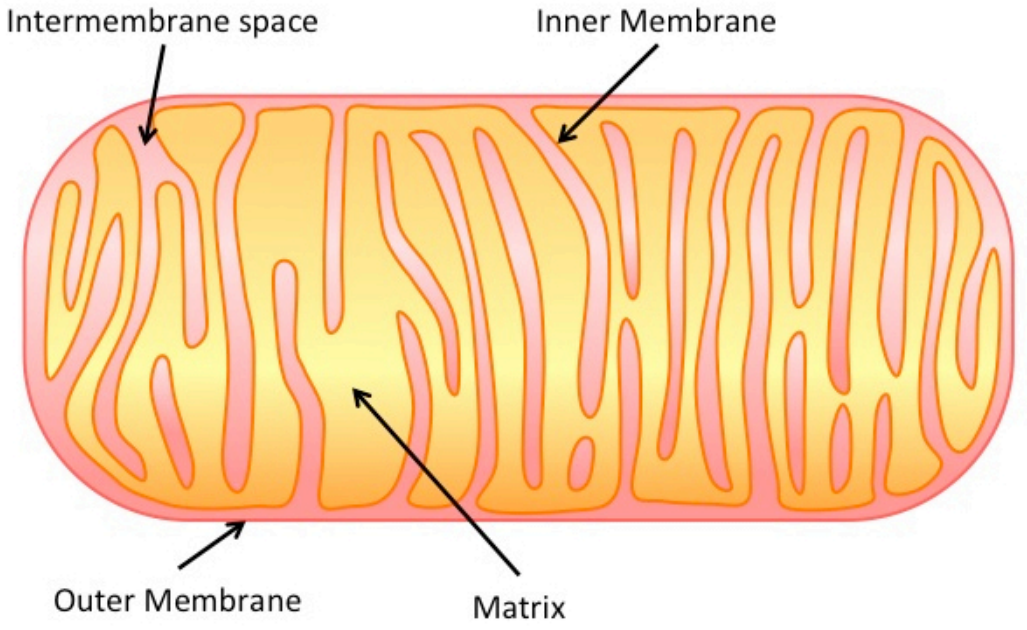


Figure 19.4 This image illustrates the structures within the mitochondria. (Image by Eva Horne and Robert Bear)

19.2 | Glycolysis

Introduction

> "My main thesis will be that in the study of the intermediate processes of metabolism we have to deal not with complex substances which elude ordinary chemical methods, but with the simple substances undergoing comprehensible reactions."

Sir Frederick Gowland Hopkins, 1933

You have read that nearly all of the energy used by living cells comes to them in the bonds of the simple 6-carbon sugar, glucose. **Glycolysis** (literally "sugar splitting") is the first step in the breakdown of glucose to extract energy for cellular metabolism. Nearly all living organisms carry out glycolysis as part of their metabolism. The process does not use oxygen and is therefore **anaerobic**. Glycolysis takes place in the cytoplasm of both prokaryotic and eukaryotic cells.

Glycolysis begins with the six carbon ring-shaped structure of a single glucose molecule and ends with two molecules of a three-carbon sugar called **pyruvate**. Glycolysis consists of two distinct phases. The first part of the glycolysis pathway traps the glucose molecule in the cell and uses energy to modify it so that the six-carbon sugar molecule can be split evenly into the two three-carbon molecules. The second part of glycolysis extracts energy from the chemical bonds in the molecules

and stores it in the form of ATP and NADH, the reduced form of NAD.

First Half of Glycolysis (Energy-Requiring Steps)

Step 1. The first step in glycolysis (**Figure 19.5**) is catalyzed by hexokinase, an enzyme with broad specificity that catalyzes the **phosphorylation** (addition of a phosphate molecule) of six-carbon sugars. Hexokinase phosphorylates glucose using ATP as the source of the phosphate, producing glucose-6-phosphate, a more high-energy form of glucose. This reaction prevents the phosphorylated glucose molecule from being transported out of the cell via glucose transporters in the plasma memrane. It can no longer leave the cell because the transport proteins recognize unmodified glucose, but not the phosphorylated version.

Step 2. In the second step of glycolysis, an isomerase converts glucose-6-phosphate into one of its isomers, fructose-6-phosphate. An **isomerase** is an enzyme that catalyzes the conversion of a molecule into one of its **isomers** (compounds with the same numbers and kinds of atoms arranged in different configurations). This change from phosphoglucose to phosphofructose facilitates the eventual split of the sugar into two three-carbon molecules.

Step 3. The third step is the phosphorylation of fructose-6-phosphate, catalyzed by the enzyme phosphofructokinase. A second ATP molecule donates a high-energy phosphate to fructose-6-phosphate, producing fructose-1,6-bisphosphate. In this pathway, phosphofructokinase is a rate-limiting enzyme. It is active when the concentration of ADP is high; it is less active when ADP levels are low and the concentration of ATP is high. Thus, if there is "sufficient" ATP in the system, the pathway slows down. This is a type of end product inhibition, since ATP is the end product of glucose catabolism.

Step 4. The newly added high-energy phosphate further destabilizes fructose-1,6-bisphosphate, which is now a very high-energy compound. The fourth step in glycolysis employs an enzyme, aldolase, to cleave fructose-1,6-bisphosphate into two phosphorylated three-carbon isomers: dihydroxyacetone-phosphate and glyceraldehyde-3-phosphate.

Step 5. In the fifth step, an isomerase transforms the dihydroxyacetone-phosphate into its isomer, glyceraldehyde-3-phosphate. Thus, the pathway will continue with two molecules of a single isomer. At this point in the pathway, there is a net investment of energy from two ATP molecules in the breakdown of one glucose molecule.

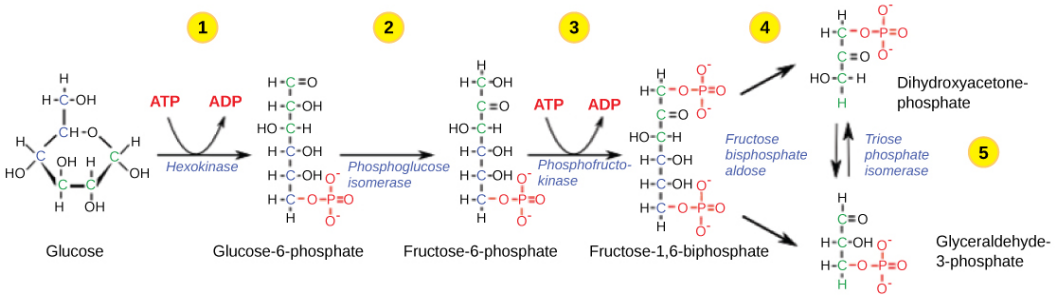


Figure 19.5 The first half of glycolysis uses two ATP molecules in the phosphorylation of glucose, which is then split into two three-carbon molecules.

Second Half of Glycolysis (Energy-Releasing Steps)

So far, glycolysis has cost the cell two ATP molecules and produced two small, three-carbon sugar molecules. Both of these molecules will proceed through the second half of the pathway, and sufficient energy will be extracted to pay back the two ATP molecules used as an initial investment and produce a profit for the cell of two additional ATP molecules and two even higher-energy NADH molecules.

Step 6. The sixth step in glycolysis (**Figure 19.6**) oxidizes the sugar (glyceraldehyde-3-phosphate), extracting high-energy electrons, which are picked up by the electron carrier NAD^+, producing NADH. The sugar is then phosphorylated by the addition of a second phosphate group, producing 1,3-bisphosphoglycerate. Note that the addition of a second phosphate group does not require another high-energy ATP molecule; inorganic phosphate ions in the cytoplasm are used instead.

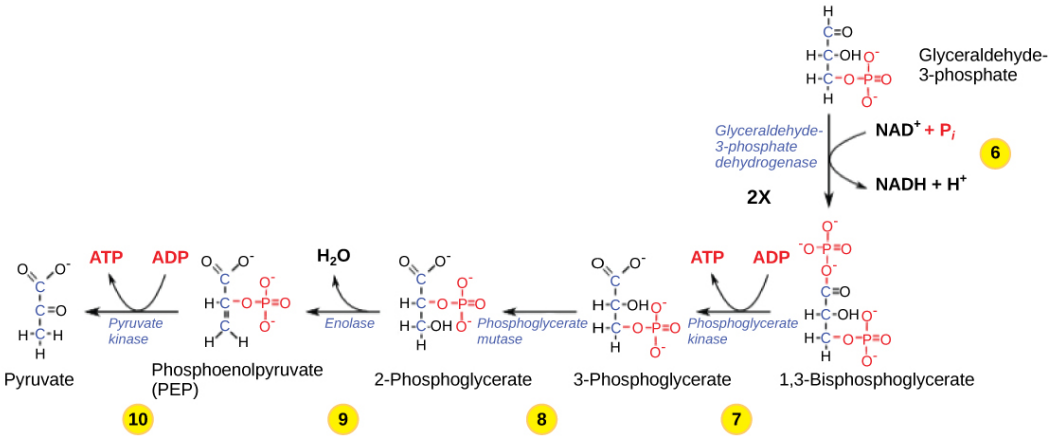


Figure 19.6 The second half of glycolysis involves phosphorylation without ATP investment (step 6) and produces two NADH and four ATP molecules per glucose.

Here again is a potential limiting factor for this pathway. The continuation of the reaction depends upon the availability of the oxidized form of the electron carrier, NAD^+. Thus, NADH must be continuously oxidized back into NAD^+ in order to keep this step going. If NAD^+ is not available, the second half of glycolysis slows down or stops. If oxygen is available in the system, the NADH will be oxidized readily, though indirectly, and the high-energy electrons from the hydrogen released in this process will be used to produce ATP. In an environment without oxygen, an alternate pathway (fermentation) can provide the oxidation of NADH to NAD^+.

Step 7. In the seventh step, catalyzed by phosphoglycerate kinase (an enzyme named for the reverse reaction), 1,3-bisphosphoglycerate donates a high-energy phosphate to ADP, forming one molecule of ATP. This is an example of **substrate-level phosphorylation**, where a phosphate group is added to ADP by removing it from another compound rather than from the phosphate ions in the cytoplasm.. A carbonyl group on the 1,3-bisphosphoglycerate is oxidized to a carboxyl group, and 3-phosphoglycerate is formed.

Step 8. In the eighth step, the remaining phosphate group in 3-phosphoglycerate moves from the third carbon to the second carbon, producing 2-phosphoglycerate (an isomer of 3-phosphoglycerate). The enzyme catalyzing this step is a mutase (isomerase).

Step 9. An enzyme called enolase catalyzes the ninth step. This enzyme causes 2-phosphoglycerate to lose water from its structure; this is a condensation reaction, resulting in the formation of a double bond that increases the potential energy in the remaining phosphate bond and produces phosphoenolpyruvate (PEP).

Step 10. The last step in glycolysis is catalyzed by the enzyme pyruvate kinase (the enzyme in this case is named for the reverse reaction of pyruvate's conversion into PEP) and results in the production of a second ATP molecule by substrate-level phosphorylation and the compound pyruvic acid (or its salt form, pyruvate). Many enzymes in enzymatic pathways are named for the reverse reactions, since the enzyme can catalyze both forward and reverse reactions (these may have been described initially by the reverse reaction that takes place *in vitro*, under non-physiological conditions).

Outcomes of Glycolysis

Glycolysis starts with glucose and produces two pyruvate molecules, a total of four ATP molecules and two molecules of NADH (**Figure 19.7**). Two ATP molecules were used in the first half of the pathway to prepare the six-carbon ring for cleavage, so the cell has a net gain of two pyruvate molecules, two ATP molecules and 2 NADH molecules for its use. If the cell cannot catabolize the pyruvate molecules further, it will harvest only these two ATP molecules from one molecule of glucose. For example, mature mammalian red blood cells are not capable of aerobic cellular respiration (they have no mitochondria), so glycolysis is their sole source of ATP.

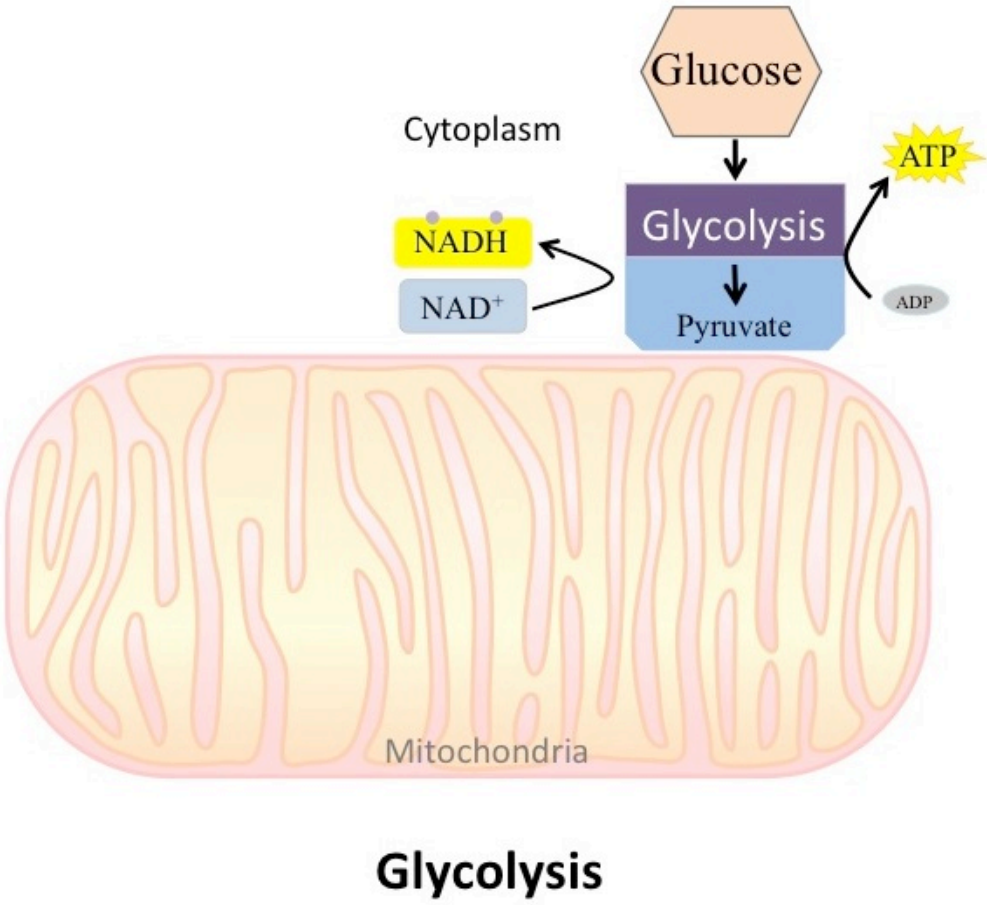


Figure 19.7 This image is an overview of glycolysis which occurs in the cytoplasm of a cell. Glucose is catabolized into 2 pyruvate molecules with 2 NAD^+ reduced to 2 NADH. Two ATP molecules are consumed, while 4 are produced, yielding a net gain of 2 ATP molecules. (Image by Eva Horne and Robert Bear

19.3 | Oxidation of Pyruvate and the Krebs Cycle

> "It is the destiny of wine to be drunk, and it is the destiny of glucose to be oxidized. But it was not oxidized immediately: its drinker kept it in his liver for more than a week, well curled up and tranquil, as a reserve aliment for a sudden effort; an effort that he was forced to make the following Sunday, pursuing a bolting horse. "

Primo Levi, *The Periodic Table*, 1975

If oxygen is available, aerobic cellular respiration will go forward. In eukaryotic cells, the pyruvate molecules produced at the end of glycolysis are transported into mitochondria, which are the sites of aerobic cellular respiration. There, pyruvate (three carbons) will be transformed into an acetyl group (two carbons) that will be attached to a carrier compound called coenzyme A (CoA). The resulting compound is called **acetyl-CoA**. CoA is made from vitamin B5, pantothenic acid. Acetyl-CoA can be used in a variety of ways by the cell, but its major function is to deliver the two-carbon energy source derived from pyruvate to the next stage of the aerobic cellular respiration pathway.

Oxidation of Pyruvate

In order for pyruvate, the product of glycolysis, to enter the next pathway, it must undergo several changes. The conversion is a three-step process (Figure 19.8).

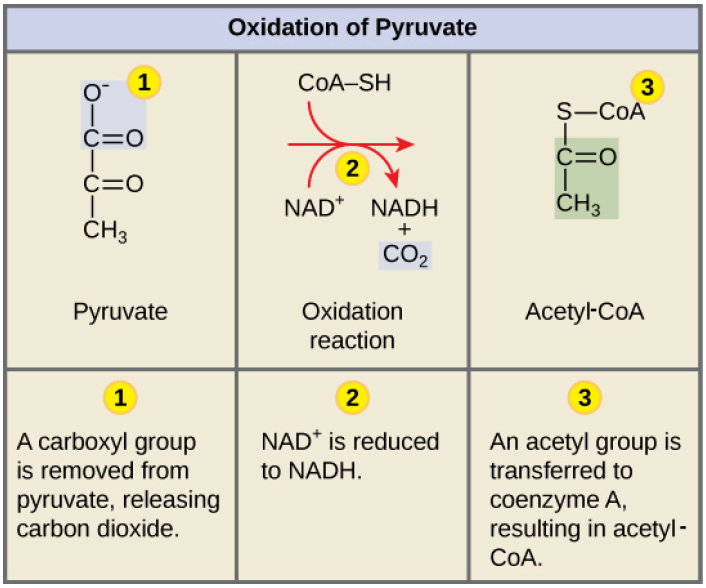


Figure 19.8 Upon entering the mitochondrial matrix, a multi-enzyme complex converts pyruvate into acetyl-CoA. In the process, carbon dioxide is released and one molecule of NADH is formed.

Step 1. A carboxyl group is removed from pyruvate, releasing a molecule of carbon dioxide into the surrounding medium. The result of this step is a two-carbon hydroxyethyl group bound to the enzyme (pyruvate dehydrogenase). This is the first of the six carbons from the original glucose molecule to be removed. This step proceeds twice for each glucose molecule (remember: there are *two* pyruvate molecules produced at the end of glycolsis). Thus, two of the six carbons will have been removed at the end of this step in aerobic cellular respiration.

Step 2. The hydroxyethyl group is oxidized to an acetyl group, and the electrons are picked up by NAD^+, forming NADH. The high-energy electrons from NADH will be used later to generate ATP.

Step 3. The enzyme-bound acetyl group is transferred to CoA, producing a molecule of acetyl-CoA.

During the oxidation of pyruvate, the incoming pyruvate is converted to acetyl-CoA, NAD^+ is reduced to NADH and a carbon dioxide is released for each pyruvate entering the stage (Figure 19.9).

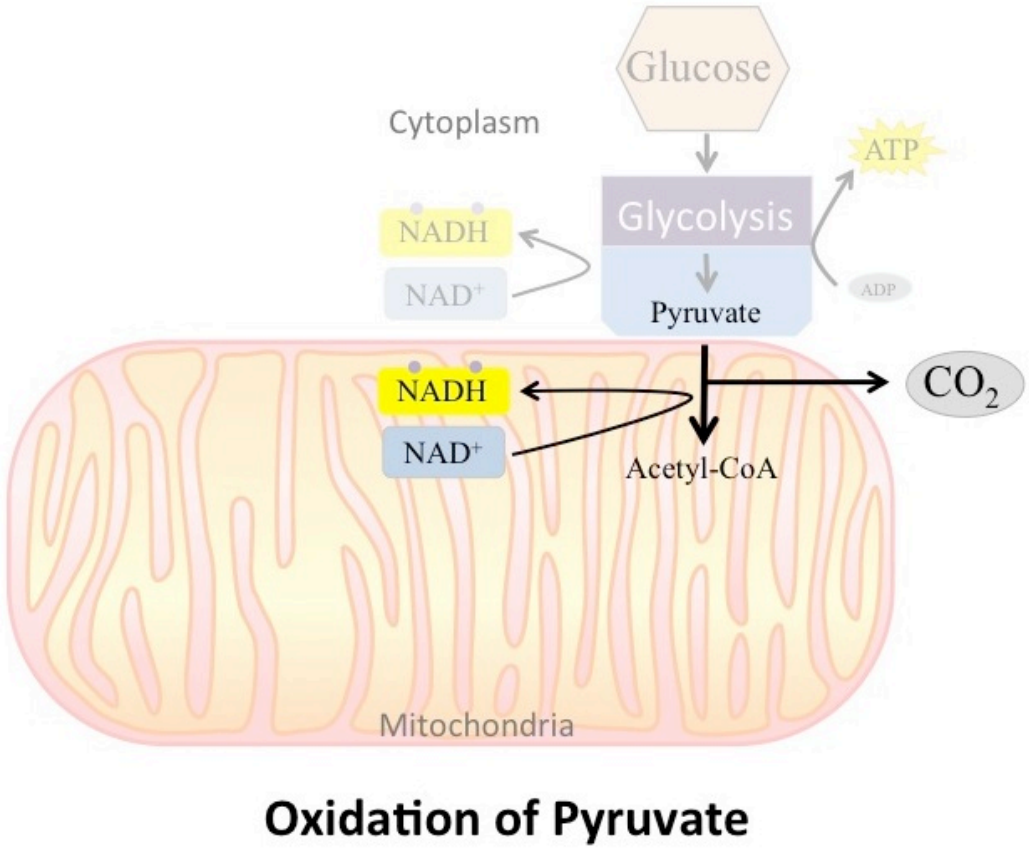


Figure 19.9 This image is an overview of the oxidation of pyruvate. During this stage, carbon dioxide is release, NAD^+ is reduced to NADH and acetyl-CoA is formed. (Image by Eva Horne and Robert Bear)

Krebs Cycle

In the presence of oxygen, the two carbons in acetyl-CoA are added to to a four-carbon molecule, oxaloacetate, to form citrate (aka citric acid), a six-carbon molecule. This is the starting point of the Krebs Cycle, which will harvest the remainder of the extractable energy from what began as a glucose molecule. This single pathway has several different names: the citric acid cycle (for citric acid, the first compound in the cycle), the tricarboxylic acid (TCA) cycle (since citric acid has 3 carboxyl groups, and is thus a tricarboxylic acids), and the **Krebs cycle**, after Hans Krebs, who first identified the steps in the pathway in the 1930s in pigeon flight muscles. This work earned Krebs a share (with Fritz Lippman) of the 1953 Nobel Prize in Physiology and Medicine.

After the conversion of pyruvate to acetyl CoA by enzymes in the mitochondrial inner membrane, the Krebs cycle takes place in the matrix of mitochondria. Almost all of the enzymes of the Krebs cycle are soluble (i.e., not bound to the membrane), with the single exception of the enzyme succinate dehydrogenase, which is embedded in the inner membrane of the mitochondrion. Unlike glycolysis, the Krebs cycle is a closed loop: The last part of the pathway regenerates the compound (oxaloacetate) used in the first step. The eight steps of the cycle are a series of redox, condensation, hydrolysis, and decarboxylation reactions that produce **two** carbon dioxide molecules, **one** GTP/ATP, and reduced forms of NADH and $FADH_2$ (**Figure 19.10**). Even though oxygen is not directly required for these reactions, this is considered an aerobic pathway because the NADH and $FADH_2$ produced must transfer their electrons to the next pathway in the system, which will use oxygen. If this transfer does not occur, there won't be any NAD or FADH regenerated, and the oxidation steps of the Krebs cycle cannot occur without those oxidized electron carriers. Note that the Krebs cycle produces very little ATP directly and does not directly consume oxygen.

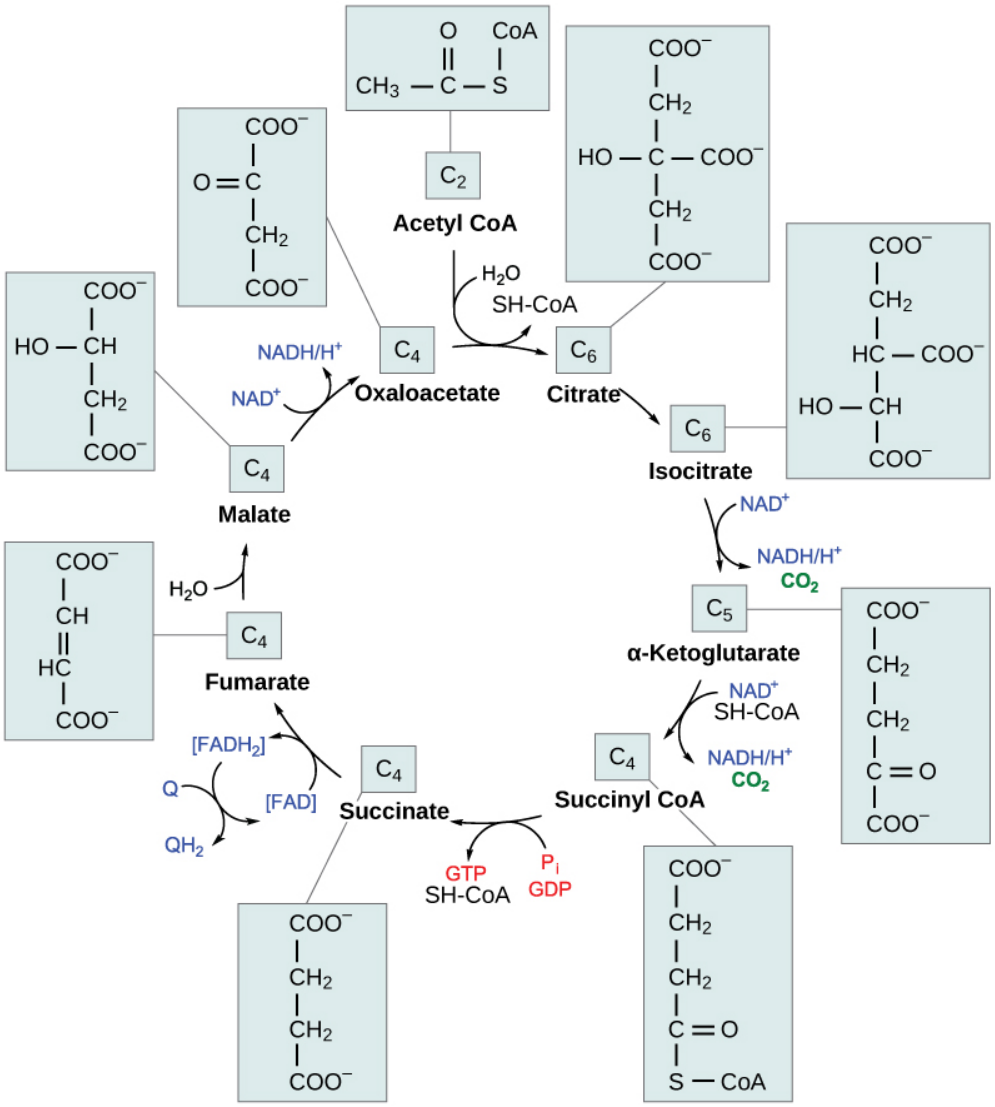


Figure 19.10 In the Krebs cycle, the acetyl group from acetyl CoA is attached to a four-carbon oxaloacetate molecule to form a six-carbon citrate molecule. Through a series of steps, citrate is oxidized, releasing two carbon dioxide molecules for each acetyl group fed into the cycle. In the process, three NAD^+ molecules are reduced to NADH, one FAD molecule is reduced to $FADH_2$, and one ATP or GTP (depending on the cell type) is produced (by substrate-level phosphorylation). Because the final product of the citric acid cycle is also the first reactant, the cycle runs continuously in the presence of sufficient reactants. (credit: modification of work by "Yikrazuul"/Wikimedia Commons)

Steps in the Krebs Cycle

Step 1. Prior to the start of the first step, a transitional phase occurs during which pyruvic acid is converted to acetyl-CoA. Then, the first step of the cycle begins: This is a condensation step, combining the two-carbon acetyl group with a four-carbon oxaloacetate molecule to form a six-carbon molecule of citrate. CoA is bound to a sulfhydryl group (-SH) and diffuses away to eventually combine with another acetyl group. This step is irreversible because it is highly exergonic. The rate of this reaction is controlled by negative feedback and the amount of ATP available. If ATP levels increase, the rate of this reaction decreases. If ATP is in short supply, the rate increases.

Step 2. In step two, citrate loses one water molecule and gains another as citrate is converted into its isomer, isocitrate.

Step 3. In step three, isocitrate is oxidized, producing a five-carbon molecule, α-ketoglutarate, together with a molecule of CO_2 and two electrons, which reduce NAD^+ to NADH. This step is also regulated by negative feedback from ATP and NADH, and a positive effect of ADP.

Steps 3 and 4. Steps three and four are both oxidation and decarboxylation steps, which release electrons that reduce NAD^+ to NADH and release carboxyl groups that form CO_2 molecules. α-Ketoglutarate is the product of step three, and a succinyl

group is the product of step four. CoA binds the succinyl group to form succinyl CoA. The enzyme that catalyzes step four is regulated by feedback inhibition of ATP, succinyl CoA, and NADH.

Step 5. In step five, a phosphate group is substituted for coenzyme A, and a high-energy bond is formed. This energy is used in substrate-level phosphorylation (during the conversion of the succinyl group to succinate) to form either guanine triphosphate (GTP) or ATP. There are two forms of the enzyme, called isoenzymes, for this step, depending upon the type of animal tissue in which they are found. One form is found in tissues that use large amounts of ATP, such as heart and skeletal muscle. This form produces ATP. The second form of the enzyme is found in tissues that have a high number of anabolic pathways, such as liver. This form produces GTP. GTP is energetically equivalent to ATP; however, its use is more restricted. In particular, protein synthesis primarily uses GTP.

Step 6. Step six is a condensation reaction that converts succinate into fumarate. Two hydrogen atoms are transferred to FAD, producing $FADH_2$. The energy contained in the electrons of these atoms is insufficient to reduce NAD^+ but adequate to reduce FAD. Unlike NADH, this carrier remains attached to the enzyme and transfers the electrons to the electron transport chain directly. This process is made possible by the localization of the enzyme catalyzing this step inside the inner membrane of the mitochondrion.

Step 7. Water is added to fumarate during step seven, and malate is produced. The last step in the Krebs cycle regenerates oxaloacetate by oxidizing malate. Another molecule of NADH is produced in the process.

Products of the Krebs Cycle

Two carbon atoms come into the Krebs cycle from each acetyl group, representing four out of the six carbons of one glucose molecule. Two carbon dioxide molecules are released on each turn of the cycle; however, these do not necessarily contain the most recently added carbon atoms. The two acetyl carbon atoms will eventually be released on later turns of the cycle; thus, all six carbon atoms from the original glucose molecule are eventually incorporated into carbon dioxide. Each turn of the cycle forms three NADH molecules and one $FADH_2$ molecule. These carriers will connect with the last portion of aerobic respiration to produce ATP molecules. One ATP is also made in each cycle (**Figure 19.11**).

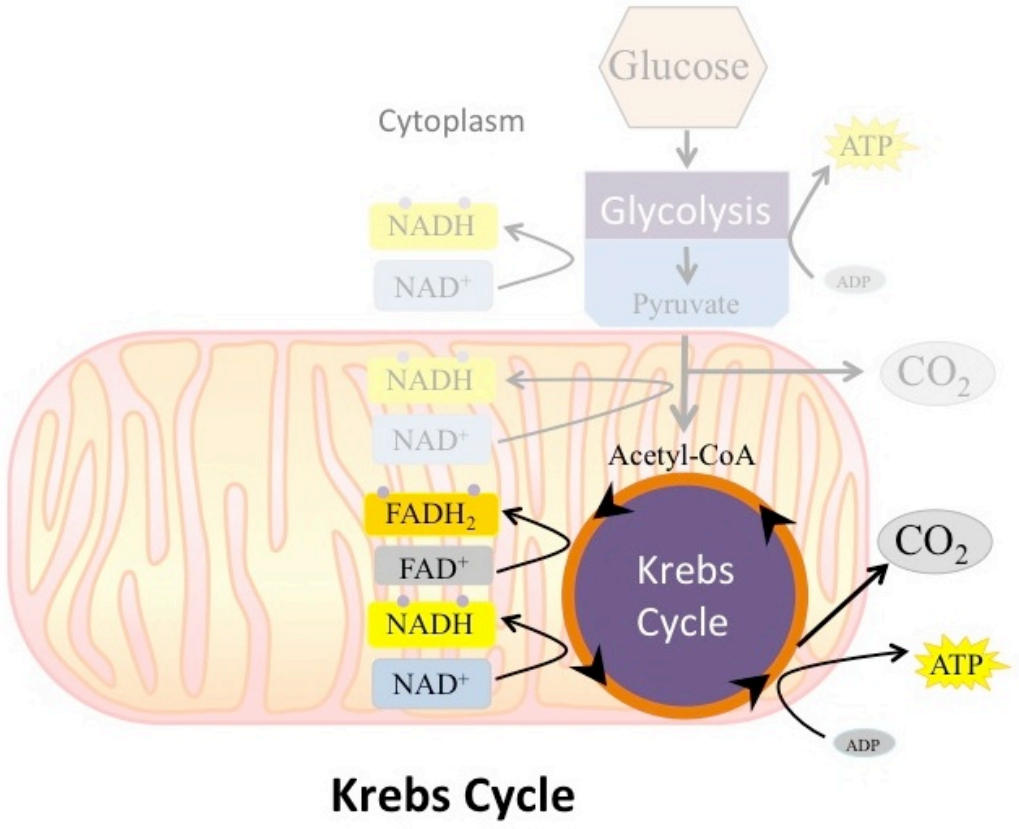


Figure 19.11 This is an overview of the Krebs cycle. (Image by Eva Horne and Robert Bear)

19.4 | Oxidative Phosphorylation

Introduction

> "Finally, to the theme of the respiratory chain, it is especially noteworthy that David Keilin's chemically simple view of the respiratory chain appears now to have been right all along — and he deserves great credit for having been so reluctant to become involved when the energy-rich chemical intermediates began to be so fashionable. This reminds me of the aphorism: "the obscure we see eventually, the completely apparent takes longer."

Peter D. Mitchell, Nobel Lecture 1978

Mitchell got the Nobel Prize for elucidating the details of one of the biggest mysteries of life — How does the energy in NADH and $FADH_2$ get converted to ATP. His explanation, now known as the **chemiosmotic hypothesis** is quite simple in concept, and involves things you already understand, like solute gradients across membranes and redox reactions. You have just read about two pathways in glucose catabolism—glycolysis and the Kreb's cycle cycle—that generate ATP. Most of the ATP generated during the aerobic catabolism of glucose, however, is not generated directly from these pathways. Rather, it is derived from a process that begins with moving electrons through a series of electron transporters that undergo redox reactions. This causes hydrogen ions to accumulate within the intermembrane space. Therefore, a concentration gradient forms; this concentration gradient represents potential energy. That energy is dissipated when hydrogen ions diffuse back into the mitochondrial matrix, passing through a channel formed by the **ATP synthase** embedded in the membrane. As they flow through, the ATP synthase uses that energy to make ATP, just like a water-powered generator uses water flow to generate electricity. The current of hydrogen ions powers the catalytic action of ATP synthase, which phosphorylates ADP, producing ATP.

Electron Transport Chain

The **electron transport chain** (Figure 19.12) is the last component of aerobic cellular respiration and is the only part of glucose metabolism that uses atmospheric oxygen. Oxygen continuously diffuses into plants and single-celled organisms; in animals, it enters the body through the respiratory system. Electron transport is a series of redox reactions that resemble a relay race or bucket brigade; electrons are passed rapidly from one component to the next, eventually arriving at the endpoint of the chain where the electrons are used to reduce molecular oxygen, producing water. There are four complexes composed of proteins, labeled I through IV in Figure 19.12, and the aggregation of these four complexes, together with associated mobile, accessory electron carriers, is called the electron transport chain. The electron transport chain is present in multiple copies in the inner mitochondrial membrane of eukaryotes and the plasma membrane of prokaryotes.

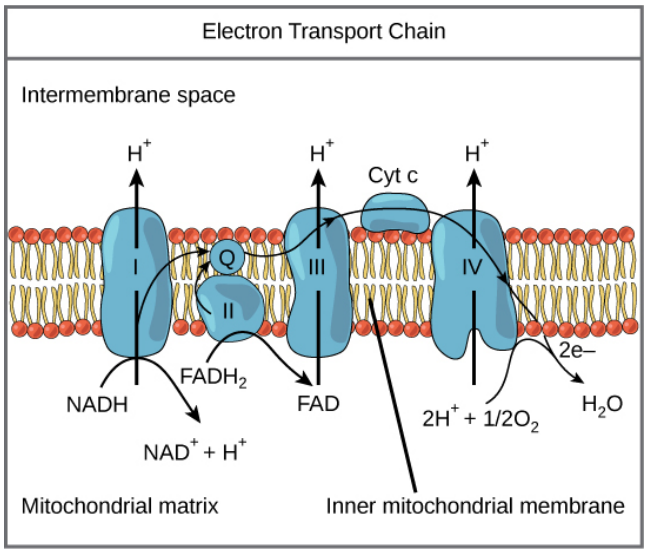


Figure 19.12 The electron transport chain is a series of electron transporters embedded in the inner mitochondrial membrane that shuttles electrons from NADH and $FADH_2$ to molecular oxygen. In the process, protons are pumped from the mitochondrial matrix to the intermembrane space, and oxygen is reduced to form water.

Complex I

To start, two electrons are carried to the first complex aboard NADH. This complex, labeled I, is composed of flavin mononucleotide (FMN) and an iron-sulfur (Fe-S)-containing protein. FMN, which is derived from vitamin B_2, also called riboflavin, is one of several prosthetic groups or co-factors in the electron transport chain. A prosthetic group is a non-protein molecule required for the activity of a protein. Prosthetic groups are organic or inorganic, non-peptide molecules bound to a protein that facilitate its function; prosthetic groups include co-enzymes, which are the prosthetic groups of enzymes. The enzyme in complex I is NADH dehydrogenase and is a very large protein, containing 45 amino acid chains. Complex I can pump four hydrogen ions across the membrane from the matrix into the intermembrane space, and it is in this way that the hydrogen ion gradient is established and maintained between the two compartments separated by the inner mitochondrial membrane.

Q and Complex II

Complex II directly receives $FADH_2$, which does not pass through complex I. The compound connecting the first and second complexes to the third is ubiquinone (Q). The Q molecule is lipid soluble and freely moves through the hydrophobic core of the membrane. Once it is reduced, (QH_2), ubiquinone delivers its electrons to the next complex in the electron transport chain. Q receives the electrons derived from NADH from complex I and the electrons derived from $FADH_2$ from complex II, including succinate dehydrogenase. This enzyme and $FADH_2$ form a small complex that delivers electrons directly to the electron transport chain, bypassing the first complex. Since these electrons bypass and thus do not energize the proton pump in the first complex, fewer ATP molecules are made from the $FADH_2$ electrons. The number of ATP molecules ultimately obtained is directly proportional to the number of protons pumped across the inner mitochondrial membrane.

Complex III

The third complex is composed of cytochrome b, another Fe-S protein, Rieske center (2Fe-2S center), and cytochrome c proteins; this complex is also called cytochrome oxidoreductase. Cytochrome proteins have a prosthetic group of heme. The heme molecule is similar to the heme in hemoglobin, but it carries electrons, not oxygen. As a result, the iron ion at its core is reduced and oxidized as it passes the electrons, fluctuating between different oxidation states: Fe^{++} (reduced) and Fe^{+++} (oxidized). The heme molecules in the cytochromes have slightly different characteristics due to the effects of the different proteins binding them, giving slightly different characteristics to each complex. Complex III pumps protons through the membrane and passes its electrons to cytochrome c for transport to the fourth complex of proteins and enzymes (cytochrome c is the acceptor of electrons from Q; however, whereas Q carries pairs of electrons, cytochrome c can accept only one at a time).

Complex IV

The fourth complex is composed of cytochrome proteins c, a, and a_3. This complex contains two heme groups (one in each of the two cytochromes, a, and a_3) and three copper ions (a pair of Cu_A and one Cu_B in cytochrome a_3). The cytochromes

hold an oxygen molecule very tightly between the iron and copper ions until the oxygen is completely reduced. The reduced oxygen then picks up two hydrogen ions from the surrounding medium to make water (H_2O). The removal of the hydrogen ions from the system contributes to the ion gradient used in the process of chemiosmosis.

Chemiosmosis

In chemiosmosis, the free energy from the series of redox reactions just described is used to pump hydrogen ions (protons) across the membrane (from inside the matrix into the intermembrane space). The uneven distribution of H^+ ions across the membrane establishes both concentration and electrical gradients (thus, an electrochemical gradient), sue to the hydrogen ions' positive charge and their aggregation on one side of the membrane. This electrochemical gradient is a form of potential energy, and its discovery was a key step in Mitchell's elucidation of the details of oxidative phosphorylation.

If the membrane was not a barrier to the movement of the hydrogen ions, the ions would diffuse back across into the matrix, driven by their electrochemical gradient. But recall that many ions cannot diffuse through the nonpolar regions of phospholipid membranes without the aid of ion channels. Similarly, hydrogen ions in the intermembrane space can only pass through the inner mitochondrial membrane through an integral membrane protein called ATP synthase (**Figure 19.13**). This complex protein acts as a tiny generator, turned by the force of the hydrogen ions diffusing down their electrochemical gradient via a channel in the protein. This powers an actual rotation of parts of the enzyme; the rotation of parts of this molecular machine facilitates the addition of a phosphate to ADP, forming ATP. In the absence of a hydrogen ion gradient, the rotation stops and no ATP is made.

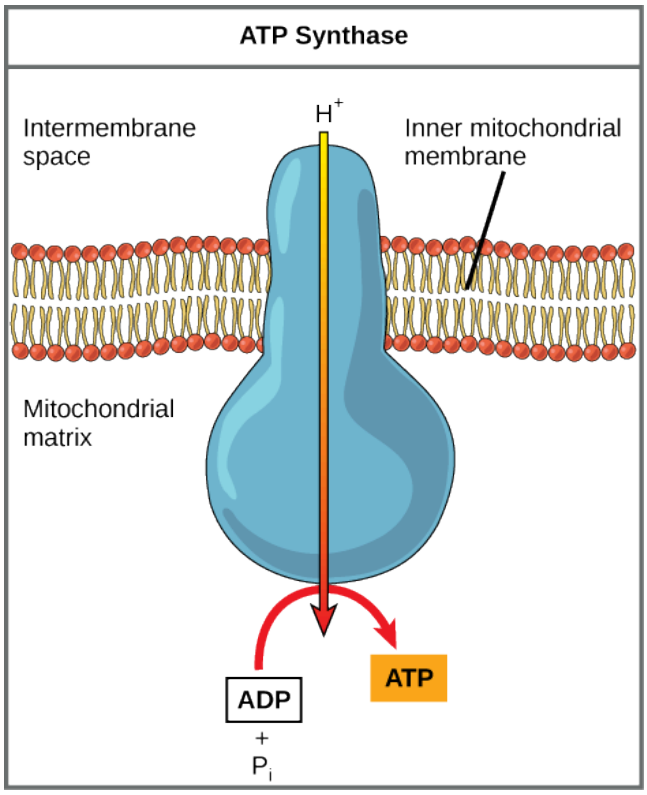


Figure 19.13 ATP synthase is a complex, molecular machine that uses a proton (H^+) gradient to form ATP from ADP and inorganic phosphate (Pi). (Credit: modification of work by Klaus Hoffmeier)

Chemiosmosis (**Figure 19.14**) is used to generate 90 percent of the ATP made during aerobic cellular respiration; it is also the method used in the light reactions of photosynthesis to harness the energy of sunlight in the process of photophosphorylation. Recall that the production of ATP using the process of chemiosmosis in mitochondria is called oxidative phosphorylation. The overall result of these reactions is the production of about 32 ATP from the energy of the electrons removed from hydrogen atoms (**Figure 19.15**). These electrons were originally part of a glucose molecule. At the end of the pathway, the electrons are used to reduce an oxygen molecule to oxygen ions. The extra electrons on the oxygen attract hydrogen ions (protons) from the surrounding medium, and water is formed.

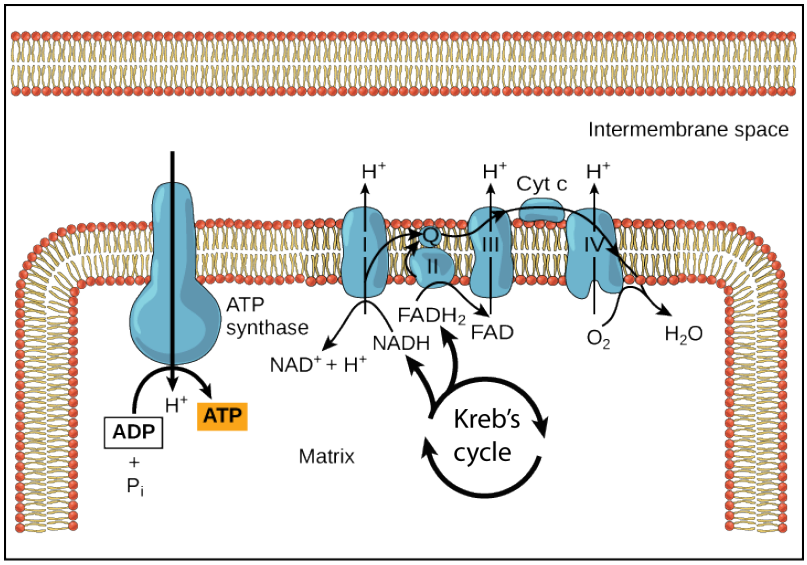


Figure 19.14 In oxidative phosphorylation, the hydrogen ion electrochemical gradient, generated by the electron transport chain, is used by ATP synthase to form ATP.

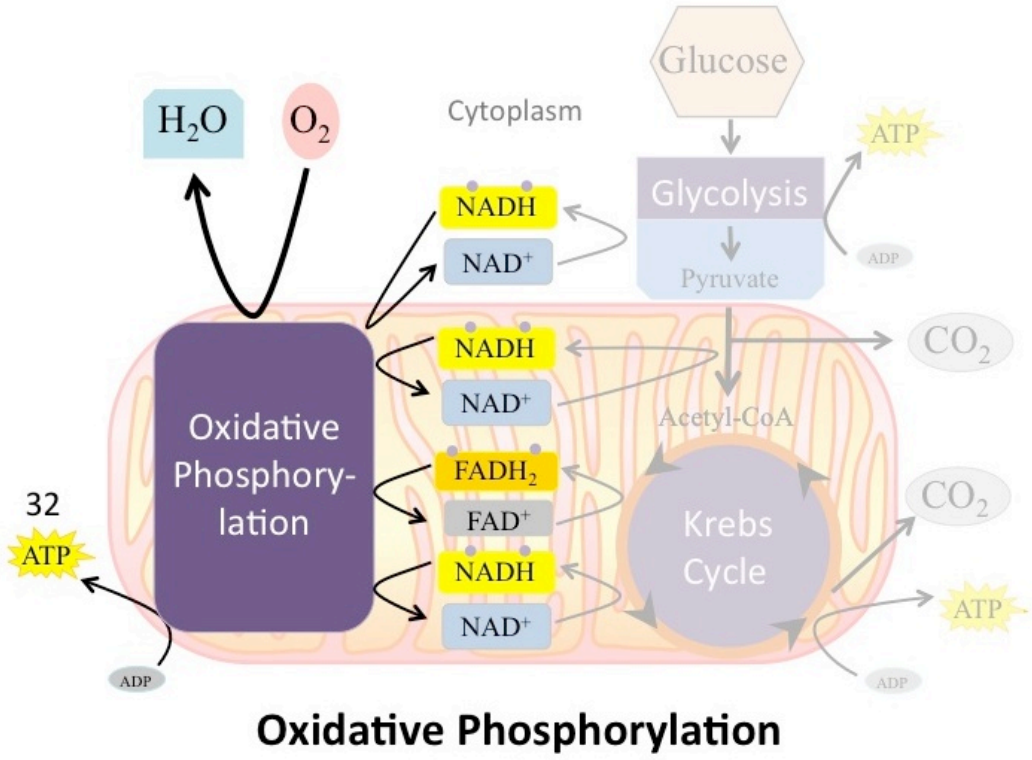


Figure 19.15 This image is an overview of oxidative phosphorylation. During this stage, the electrons from NADH and $FADH^2$ are used to produce a large amount of ATP (~32). (Image by Eva Horne and Robert Bear)

ATP Yield

The number of ATP molecules generated from the catabolism of glucose can be variable. For example, the number of hydrogen ions that the electron transport chain complexes can pump through the membrane varies between species. Another source of variance stems from the shuttle of electrons across the membranes of the mitochondria. (The NADH generated from glycolysis cannot easily enter mitochondria.) Thus, electrons are picked up on the inside of mitochondria by either NAD^+ or FAD^+. As you have learned earlier, these FAD^+ molecules carry electrons that are lower in energy than those

on NADH; consequently, fewer ATP molecules are generated when FAD^+ acts as a carrier. NAD^+ is used as the electron transporter in the liver and FAD^+ acts in the brain. Assuming optimal ATP production, a net of 2 are formed in glycolysis, 2 are formed in the Kreb's cycle, and 32 produced during oxidative phophorylation (**Figure 19.16**). The total amount of ATP formed form 1 molecule of glucose is generally around 36 under optimal conditions.

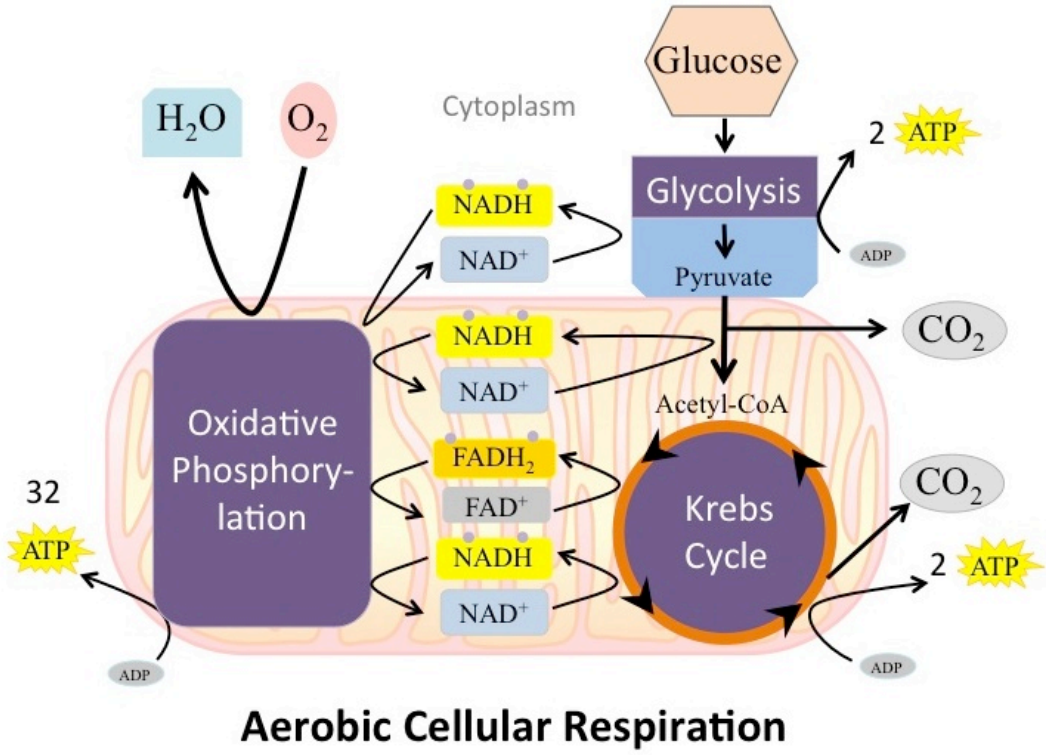


Figure 19.16 This image is an overview of aerobic cellular respiration showing the number of ATP produced by the various steps of the process. Assuming optimal conditions, there are 36 ATP molecules produced for each molecule of glucose entering aerobic cellular respiration. (Image by Eva Horne and Robert Bear)

Another factor that affects the yield of ATP molecules generated from glucose is the fact that intermediate compounds and electron carriers in these pathways are used for other purposes. Glucose catabolism connects with the pathways that build or break down all other biochemical compounds in cells, and the result is that some metabolites (e.g. pyruvate) are shuttled off for other purposes. Additionally, sugars other than glucose are fed into various steps in the glycolytic pathway for energy extraction. The five-carbon sugars that form nucleic acids are made from intermediates in glycolysis. Certain nonessential amino acids can be made from intermediates of both glycolysis and the Krebs cycle. Lipids, such as cholesterol and triglycerides, are also made from acetyl-CoA, and both amino acids and triglycerides are broken down for energy, feeding into these pathways at various places. NADH is used by other redox reactions in cells, which means that not all of the high energy electrons liberated from glucose are used in oxidative phosphorylation. So the tidy pictures above don't really reflect the complexity of metabolism in a living cell. So the relationship between glucose breakdown and ATP production is like all relationships: "It's complicated."

Codependency of Photosynthesis and Aerobic Cellular Respiration

As you may have noticed, the substrates of aerobic cellular respiration are the products of photosynthesis, and the substrates of photosynthesis are the products of aerobic cellular respiration (**Figure 19.17**). These two processes are interdependent; they have been evolving with one another for a very long time and are important components of the carbon cycle and obviously the flow of energy. In this section, we will take a closer look at the codependency of photosynthesis and cellular respiration.

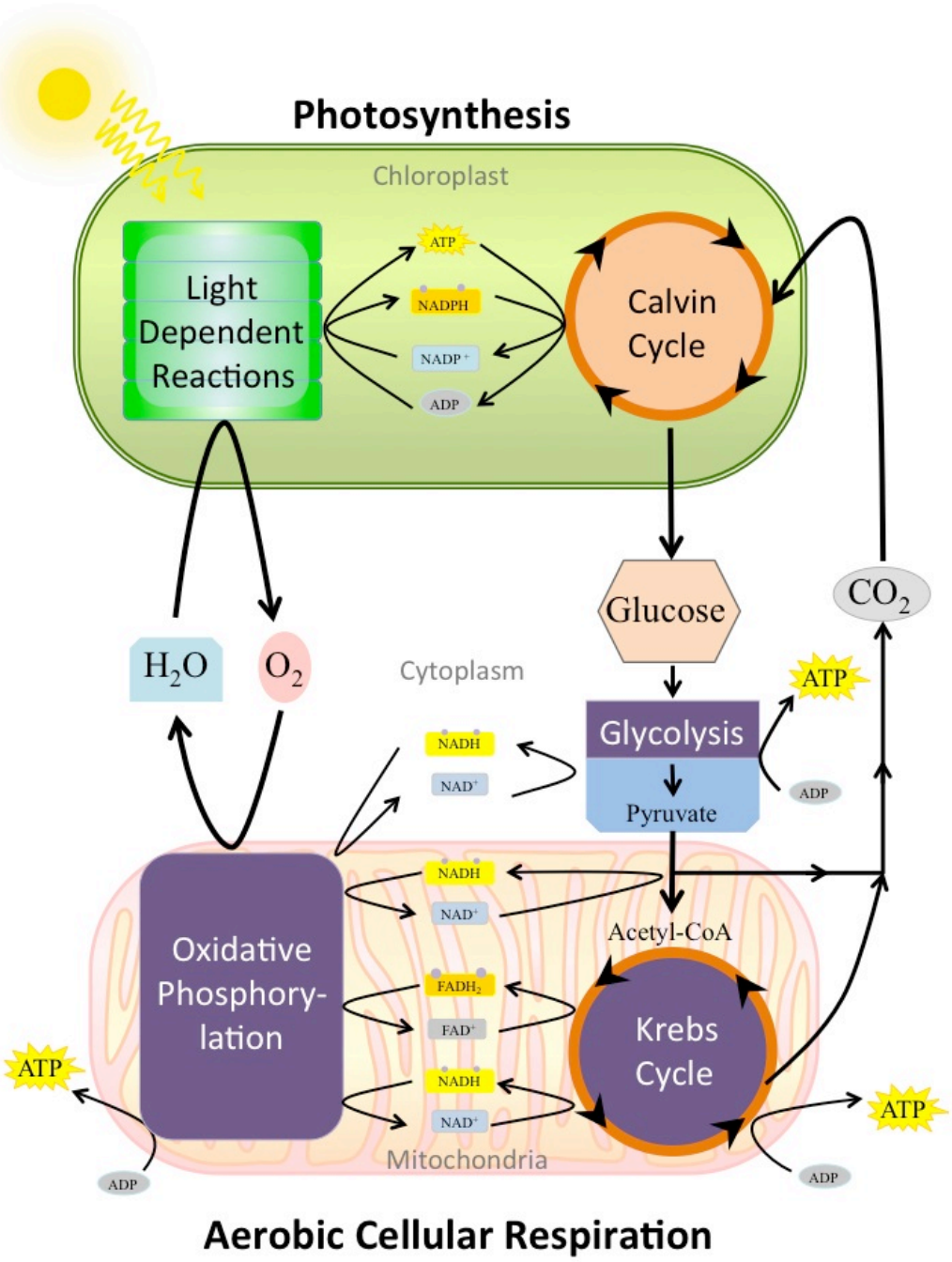


Figure 19.17 This image shows the interrelatedness of photosynthesis and aerobic cellular respiration Note how the the two processes are almost mirror images of each other. (Image by Eva Horne and Robert Bear)

The first point of interaction is the cycling of water and oxygen. During photosynthesis, oxygen gas is produced and the source of the oxygen atoms is the reactant water. The oxygen is formed in the initial step of photosynthesis, when photosystem II splits a water molecule into oxygen gas, hydrogen ions and electrons. In the light dependent reactions on the thylakoid membrane, light energy is absorbed and transferred to ATP and NADP. The light energy is transferred to chemical energy during this process. In aerobic cellular respiration, the roles are reversed: oxygen is the reactant, and water is the product. The aerobic cellular respiration process uses oxygen in the final (not initial) step, when it merges (rather than splitting) with the electrons and hydrogen ions during oxidative phosphorylation. The cycling of water and oxygen between photosynthesis and aerobic cellular respiration is the source of electrons that harvest, store and transfer energy.

Another product of photosynthesis is Glucose (carbohydrate). In the Calvin cycle, carbohydrates are produced by fixing

carbon dioxide (reactant) and by adding energy from ATP and NADPH. During aerobic cellular respiration, NADH and $FADH_2$ capture the energy stored in glucose (reactant). The captured energy in NADH and $FADH_2$ is then used during oxidative phospholrylation to produce ATP. This process of energy capture also releases a waste product - carbon dioxide. So energy from the sun is used to make glucose from CO_2, which is a nutrient for the plant. The product of photosynthesis (glucose) contains that nutrient AND the energy derived from sunlight; glcuose is thus a high energy nutrient for heterotrophic organisms. These organisms release the nutrient as CO_2 and use some of the energy to do cellular work. Other energy is lost as heat. It should be obvious that nutrients cycle between autotrophs and heterotrophs, while energy moves in a one-way direction from autotrophs to heterotrophs.

Photosynthesis and aerobic cellular respiration are codependent; they rely upon use each other's products and byproducts and are essentially the opposites of each other. Neither process can happen without the other because if there is no carbon dioxide, photosynthesis cannot proceed; and if there is no oxygen, aerobic cellular respiration cannot proceed. The exergonic reactions of cellular respiration (glycolysis, Kreb's cycle and oxidative phosphorylation) mirror the energonic reactions of photosynthesis (Light Dependent Reactions and Calvin cycle). As we learned on the ecosystem level, nutrients cycle and energy flows. The same is true on the cellular level: nutrients cycle and energy flows.

19.5 | Metabolism Without Oxygen

> "Not all chemicals are bad. Without chemicals such as hydrogen and oxygen, for example, there would be no way to make water, a vital ingredient in beer."

- Dave Barry

The other vital ingredient in beer, of course, is ethanol. And ethanol is one of the more important products of anaerobic respiration. Other compounds, such as lactic acid, are also metabolic by-products of respiration in the absence of oxygen. The key purpose of all of these metabolic pathways is the need to regenerate NAD from the NADH produced during glycolysis. In aerobic respiration, NADH donates those electrons to the electron transfer chain. Recall that the final electron acceptor in that pathway is an oxygen molecule, O_2. If oxygen is present, then ATP will be produced using the energy of high-energy electrons carried by NADH or $FADH_2$ to the electron transport chain. If oxygen is not present, NADH must be reoxidized to NAD^+ for reuse as an electron carrier for the glycolytic pathway to continue. How is this done? Some living systems use an organic molecule as the final electron acceptor. Processes that use an organic molecule to regenerate NAD^+ from NADH are collectively referred to as **fermentation** (anaerobic metabolism). In contrast, some living systems use an inorganic molecule as a final electron acceptor of the electron transport chain, and this process is called **anaerobic cellular respiration** in which organisms convert energy for their use in the absence of oxygen.

Anaerobic Metabolism

Certain prokaryotes, including some species of bacteria and Archaea, use anaerobic respiration. For example, the group of Archaea called methanogens reduces carbon dioxide to methane to oxidize NADH. These microorganisms are found in soil and in the digestive tracts of ruminants, such as cows and sheep. Similarly, sulfate-reducing bacteria and Archaea, most of which are anaerobic (**Figure 19.18**), reduce sulfate to hydrogen sulfide to regenerate NAD^+ from NADH.

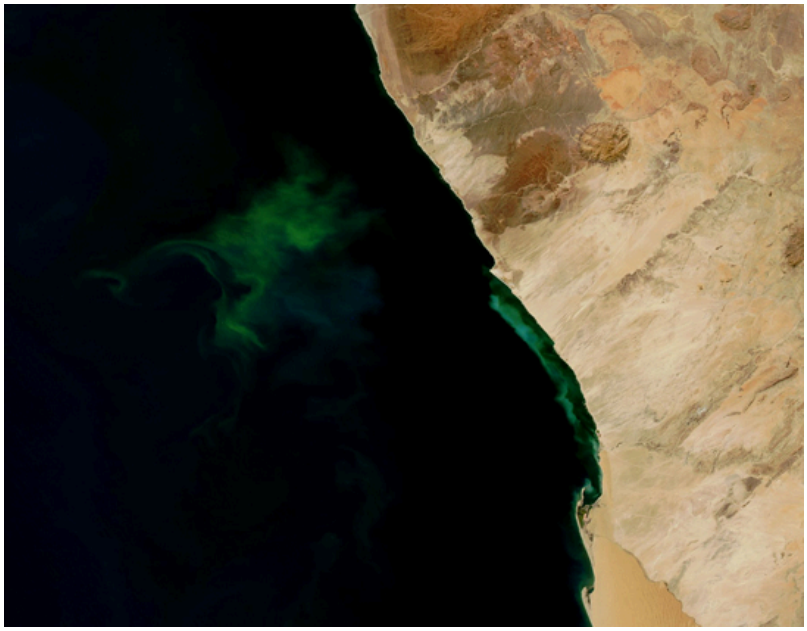


Figure 19.18 The green color seen in these coastal waters is from an eruption of hydrogen sulfide-producing bacteria. These anaerobic, sulfate-reducing bacteria release hydrogen sulfide gas as they decompose algae in the water. (credit: modification of work by NASA/Jeff Schmaltz, MODIS Land Rapid Response Team at NASA GSFC, Visible Earth Catalog of NASA images)

Lactic Acid Fermentation

The fermentation method used by animals and certain bacteria, like those in yogurt, is lactic acid fermentation (**Figure 19.19**). This type of fermentation is used routinely in mammalian red blood cells and in skeletal muscle that has an insufficient oxygen supply to allow aerobic respiration to continue (that is, in muscles used to the point of fatigue). In muscles, lactic acid accumulation must be removed by the blood circulation and the lactate brought to the liver for further metabolism.

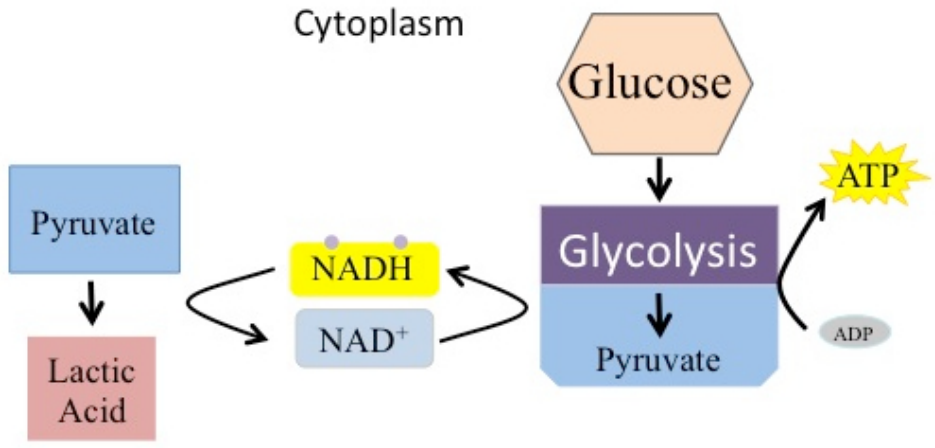


Figure 19.19 Lactic acid fermentation is common in muscle cells that have run out of oxygen. (Image by Robert Bear)

Alcohol Fermentation

Another familiar fermentation process is alcohol fermentation (**Figure 19.20**) that produces ethanol, an alcohol. The first chemical reaction of alcohol fermentation is the following (CO_2 does not participate in the second reaction):

The first reaction is catalyzed by pyruvate decarboxylase, a cytoplasmic enzyme, with a coenzyme of thiamine pyrophosphate (TPP, derived from vitamin B_1 and also called thiamine). A carboxyl group is removed from pyruvate, releasing carbon dioxide as a gas. The loss of carbon dioxide reduces the size of the molecule by one carbon, making

acetaldehyde. The second reaction is catalyzed by alcohol dehydrogenase to oxidize NADH to NAD^+ and reduce acetaldehyde to ethanol. The fermentation of pyruvate by yeast produces the ethanol found in alcoholic beverages. Ethanol tolerance of yeast is variable, ranging from about 5 percent to 21 percent, depending on the yeast strain and environmental conditions.

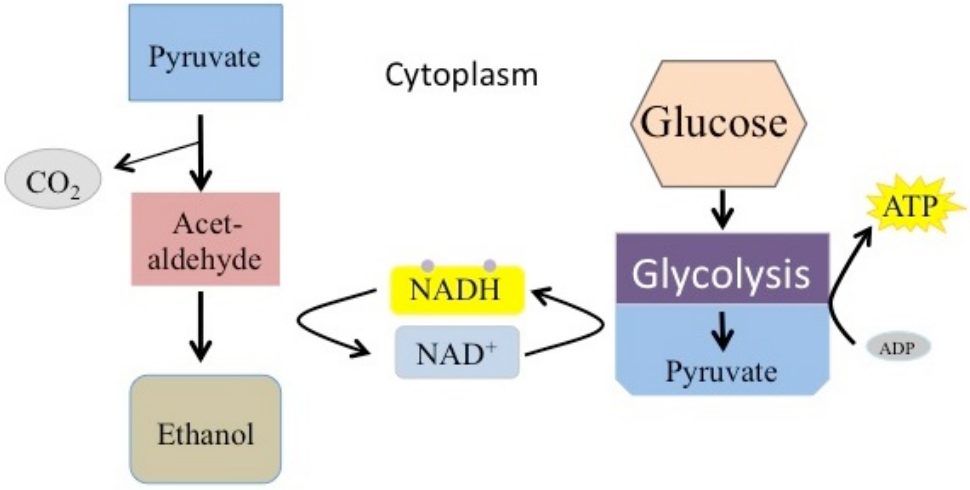


Figure 19.20 Alcohol fermentation is occurs in fungi and plants with carbon dioxide and alcohol as the byproducts. (Image by Robert Bear)

Other Types of Fermentation

Other fermentation methods occur in bacteria. Many prokaryotes are facultatively anaerobic. This means that they can switch between aerobic respiration and fermentation, depending on the availability of oxygen. Certain prokaryotes, like *Clostridia*, are obligate anaerobes. Obligate anaerobes live and grow in the absence of molecular oxygen. Oxygen is a poison to these microorganisms and kills them on exposure. It should be noted that all forms of fermentation, except lactic acid fermentation, produce gas. The production of particular types of gas is used as an indicator of the fermentation of specific carbohydrates, which plays a role in the laboratory identification of the bacteria. Various methods of fermentation are used by assorted organisms to ensure an adequate supply of NAD^+ for the sixth step in glycolysis. Without these pathways, that step would not occur and no ATP would be harvested from the breakdown of glucose.

20 | BACTERIA AND FUNGI: USING ALTERNATIVE ENERGY SOURCES

20.1 | Connections of Carbohydrate, Protein, and Lipid Metabolic Pathways

Introduction

> “ SIR TOBY: Does not our lives consist of the four elements? / SIR ANDREW: Faith, so they say; but I think it rather consists of eating and drinking. / SIR TOBY: Thou'rt a scholar; let us therefore eat and drink. ”

William Shakespeare, *Twelfth Night*

You have learned about the catabolism of glucose, which provides energy to living cells. But when you eat and drink, as advised by Sir Toby, you consume more than glucose for food. How does a turkey/avocado sandwich end up as ATP in your cells? This happens because all of the catabolic pathways for carbohydrates, proteins, and lipids eventually connect into glycolysis and the Krebs cycle pathways (see **Figure 20.2**). Metabolic pathways should be thought of as porous—that is, substances enter from other pathways, and intermediates leave for other pathways. These pathways are not closed systems. Many of the substrates, intermediates, and products in a particular pathway are reactants in other pathways.

Connections of Other Sugars to Glucose Metabolism

Glycogen, a polymer of glucose, is an energy storage molecule in animals. When there is adequate ATP present, excess glucose is shunted into glycogen for storage. Glycogen is made and stored in both liver and muscle. The glycogen will be hydrolyzed into glucose monomers (G-1-P) if blood sugar levels drop. The presence of glycogen as a source of glucose allows ATP to be produced for a longer period of time during exercise. Glycogen is broken down into G-1-P and converted into G-6-P in both muscle and liver cells, and this product enters the glycolytic pathway.

Sucrose is a disaccharide with a molecule of glucose and a molecule of fructose bonded together with a glycosidic linkage. Fructose is one of the three dietary monosaccharides, along with glucose and galactose (which is part of the milk sugar, the disaccharide lactose), which are absorbed directly into the bloodstream during digestion. The catabolism of both fructose and galactose produces the same number of ATP molecules as glucose.

Connections of Proteins to Glucose Metabolism

Proteins are hydrolyzed by a variety of enzymes in cells. Most of the time, the amino acids are recycled into the synthesis of new proteins. If there are excess amino acids, however, or if the body is in a state of starvation, some amino acids will be shunted into the pathways of glucose catabolism (**Figure 20.1**). Each amino acid must have its amino group removed prior to entry into these pathways. The amino group is converted into ammonia. In mammals, the liver synthesizes urea from two

ammonia molecules and a carbon dioxide molecule. Thus, urea is the principal waste product in mammals produced from the nitrogen originating in amino acids, and it leaves the body in urine.

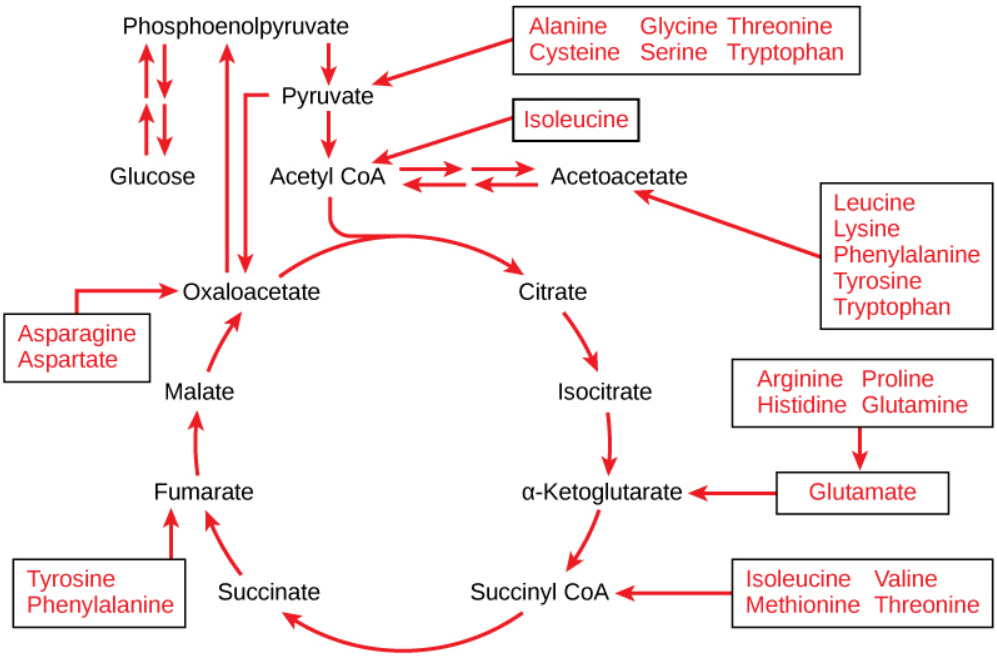


Figure 20.1 The carbon skeletons of certain amino acids (indicated in boxes) derived from proteins can feed into the Krebs cycle. (credit: modification of work by Mikael Häggström)

Connections of Lipid and Glucose Metabolisms

The lipids that are connected to the glucose pathways are cholesterol and triglycerides. Cholesterol is a lipid that contributes to plasma membrane flexibility and is a precursor of steroid hormones. The synthesis of cholesterol starts with acetyl groups and proceeds in only one direction. The process cannot be reversed.

Triglycerides are a form of long-term energy storage in animals. Triglycerides are made of glycerol and three fatty acids. Animals can make most of the fatty acids they need. Triglycerides can be both made and broken down through parts of the glucose catabolism pathways. Glycerol can be phosphorylated to glycerol-3-phosphate, which continues through glycolysis. Fatty acids are catabolized in a process called beta-oxidation that takes place in the matrix of the mitochondria and converts their fatty acid chains into two carbon units of acetyl groups. The acetyl groups are picked up by CoA to form acetyl-CoA that proceeds into the Krebs cycle.

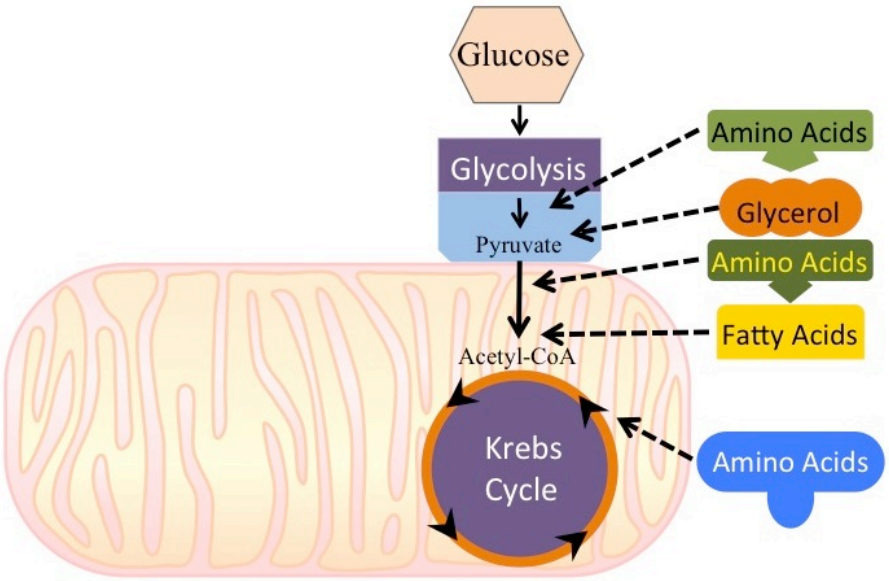


Figure 20.2 Glycogen from the liver and muscles, hydrolyzed into glucose-1-phosphate, together with fats and proteins, can feed into the catabolic pathways for carbohydrates. Work by Eva Horne and Robert A. Bear.

evolution CONNECTION

Pathways of Photosynthesis and Cellular Metabolism

The processes of photosynthesis and cellular metabolism consist of several very complex pathways. It is generally thought that the first cells arose in an aqueous environment—a "soup" of nutrients—probably on the surface of some porous clays. If these cells reproduced successfully and their numbers climbed steadily, it follows that the cells would begin to deplete the nutrients from the medium in which they lived as they shifted the nutrients into the components of their own bodies. This hypothetical situation would have resulted in natural selection favoring those organisms that could exist by using the nutrients that remained in their environment and by manipulating these nutrients into materials upon which they could survive. Selection would favor those organisms that could extract maximal value from the nutrients to which they had access.

An early form of photosynthesis developed that harnessed the sun's energy did not use water as a source of hydrogen atoms, and did not produce free oxygen (anoxygenic photosynthesis). Early photosynthesis did not produce free oxygen because it did not use water as the source of hydrogen ions; instead, it used materials like hydrogen sulfide and consequently produced sulfur. It is thought that glycolysis developed at this time and could take advantage of the simple sugars being produced, but these reactions were unable to fully extract the energy stored in the carbohydrates. The development of glycolysis probably predated the evolution of photosynthesis, as it was well suited to extract energy from materials spontaneously accumulating in the "primeval soup." A later form of photosynthesis used water as a source of electrons and hydrogen, and generated free oxygen. Over time, the atmosphere became oxygenated, but not before the oxygen released oxidized metals in the ocean and created a "rust" layer in the sediment, permitting the dating of the rise of the first oxygenic photosynthesizers. Living things adapted to exploit this new atmosphere that allowed aerobic respiration as we know it to evolve. When the full process of oxygenic photosynthesis developed and the atmosphere became oxygenated, cells were finally able to use the oxygen expelled by photosynthesis to extract considerably more energy from the sugar molecules using the Krebs cycle and oxidative phosphorylation.

20.2 | Prokaryotes

Introduction

> "Perhaps bacteria may tentatively be regarded as biochemical experiments; owing to their relatively small size and rapid growth, variations must arise much more frequently than in more differentiated forms of life, and they can in addition afford to occupy more precarious positions in natural economy than larger organisms with more exacting requirements."

Marjory Stephenson, in *Bacterial Metabolism*, (1930)

Prokaryotic Diversity

Prokaryotes are ubiquitous, and, as mentioned above, highly diverse in their metabolic activities. They cover every imaginable surface where there is sufficient moisture, and they live on and inside of other living things. In the typical human body, prokaryotic cells outnumber human body cells by about ten to one. They comprise the majority of living things in all ecosystems. Some prokaryotes thrive in environments that are inhospitable for most living things. Prokaryotes recycle **nutrients**—essential substances (such as carbon and nitrogen)—and they drive the evolution of new ecosystems, some of which are natural and others man-made. Prokaryotes have been on Earth since long before multicellular life appeared.

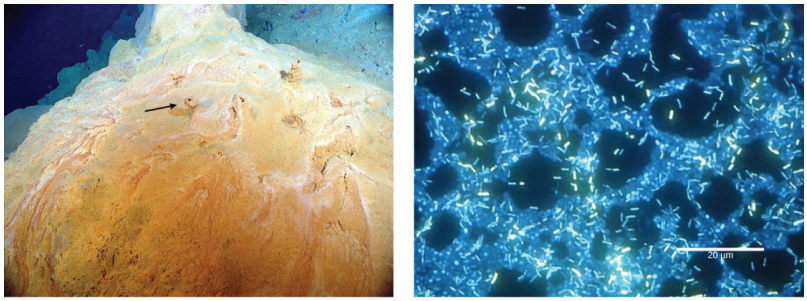
Prokaryotes, the First Inhabitants of Earth

When and where did life begin? What were the conditions on Earth when life began? Prokaryotes were the first forms of life on Earth, and they existed for billions of years before plants and animals appeared. The Earth and its moon are thought to be about 4.54 billion years old. This estimate is based on evidence from radiometric dating of meteorite material together with other substrate material from Earth and the moon. Early Earth had a very different atmosphere (contained less molecular oxygen) than it does today and was subjected to strong radiation; thus, the first organisms would have flourished where they were more protected, such as in ocean depths or beneath the surface of the Earth. At this time too, strong volcanic activity was common on Earth, so it is likely that these first organisms—the first prokaryotes—were adapted to very high temperatures. Early Earth was prone to geological upheaval and volcanic eruption, and was subject to bombardment by mutagenic radiation from the sun. The first organisms were prokaryotes that could withstand these harsh conditions.

Microbial Mats

Microbial mats or large biofilms may represent the earliest forms of life on Earth; there is fossil evidence of their presence starting about 3.5 billion years ago. A **microbial mat** is a multi-layered sheet of prokaryotes (**Figure 20.3**) that includes mostly bacteria, but also archaea. Microbial mats are a few centimeters thick, and they typically grow where different types of materials interface, mostly on moist surfaces. The various types of prokaryotes that comprise them carry out different metabolic pathways, and that is the reason for their various colors. Prokaryotes in a microbial mat are held together by a glue-like sticky substance that they secrete called extracellular matrix.

The first microbial mats likely obtained their energy from chemicals found near hydrothermal vents. A **hydrothermal vent** is a breakage or fissure in the Earth's surface that releases geothermally heated water. With the evolution of photosynthesis about 3 billion years ago, some prokaryotes in microbial mats came to use a more widely available energy source—sunlight—whereas others were still dependent on chemicals from hydrothermal vents for energy and food.



(a) (b)

Figure 20.3 This (a) microbial mat, about one meter in diameter, grows over a hydrothermal vent in the Pacific Ocean in a region known as the "Pacific Ring of Fire." The mat helps retain microbial nutrients. Chimneys such as the one indicated by the arrow allow gases to escape. (b) In this micrograph, bacteria are visualized using fluorescence microscopy. (credit a: modification of work by Dr. Bob Embley, NOAA PMEL, Chief Scientist; credit b: modification of work by Ricardo Murga, Rodney Donlan, CDC; scale-bar data from Matt Russell)

Stromatolites

Fossilized microbial mats represent the earliest record of life on Earth. A **stromatolite** is a sedimentary structure formed when minerals precipitate out of water by prokaryotes in a microbial mat (**Figure 20.4**). Stromatolites form layered rocks made of carbonate or silicate. Although most stromatolites are artifacts from the past, there are places on Earth where stromatolites are still forming. For example, growing stromatolites have been found in the Anza-Borrego Desert State Park in San Diego County, California.

(a) (b)

Figure 20.4 (a) These living stromatolites are located in Shark Bay, Australia. (b) These fossilized stromatolites, found in Glacier National Park, Montana, are nearly 1.5 billion years old. (credit a: Robert Young; credit b: P. Carrara, NPS)

The Ancient Atmosphere

Evidence indicates that during the first two billion years of Earth's existence, the atmosphere was **anoxic**, meaning that there was no molecular oxygen. Therefore, only those organisms that can grow without oxygen— **anaerobic** organisms—were able to live. Autotrophic organisms that convert solar energy into chemical energy are called **phototrophs**, and they appeared within one billion years of the formation of Earth. Then, **cyanobacteria**, also known as blue-green algae, evolved from these simple phototrophs one billion years later. Cyanobacteria (**Figure 20.5**) began the oxygenation of the atmosphere. Increased atmospheric oxygen allowed the development of more efficient O_2-utilizing catabolic pathways. It also opened up the land to increased colonization, because some O_2 is converted into O_3 (ozone) and ozone effectively absorbs the ultraviolet light that would otherwise cause lethal mutations in DNA. Ultimately, the increase in O_2 concentrations allowed the evolution of other life forms.

Figure 20.5 This hot spring in Yellowstone National Park flows toward the foreground. Cyanobacteria in the spring are green, and as water flows down the gradient, the intensity of the color increases as cell density increases. The water is cooler at the edges of the stream than in the center, causing the edges to appear greener. (credit: Graciela Brelles-Mariño)

Microbes Are Adaptable: Life in Moderate and Extreme Environments

Some organisms have developed strategies that allow them to survive harsh conditions. Prokaryotes thrive in a vast array of environments: Some grow in conditions that would seem very normal to us, whereas others are able to thrive and grow under conditions that would kill a plant or animal. Almost all prokaryotes have a cell wall, a protective structure that allows them to survive in both hyper- and hypo-osmotic conditions. Some soil bacteria are able to form endospores that resist heat and drought, thereby allowing the organism to survive until favorable conditions recur. These adaptations, along with others, allow bacteria to be the most abundant life form in all terrestrial and aquatic ecosystems.

Other bacteria and archaea are adapted to grow under extreme conditions and are called **extremophiles**, meaning "lovers of extremes." Extremophiles have been found in all kinds of environments: the depth of the oceans, hot springs, the Artic and the Antarctic, in very dry places, deep inside Earth, in harsh chemical environments, and in high radiation environments (**Figure 20.6**), just to mention a few. These organisms give us a better understanding of prokaryotic diversity and open up the possibility of finding new prokaryotic species that may lead to the discovery of new therapeutic drugs or have industrial applications. Because they have specialized adaptations that allow them to live in extreme conditions, many extremophiles cannot survive in moderate environments. There are many different groups of extremophiles: They are identified based on the conditions in which they grow best, and several habitats are extreme in multiple ways. For example, a soda lake is both salty and alkaline, so organisms that live in a soda lake must be both alkaliphiles and halophiles (**Table 20.1**). Other extremophiles, like **radioresistant** organisms, do not prefer an extreme environment (in this case, one with high levels of radiation), but have adapted to survive in it (**Figure 20.6**).

Extremophiles and Their Preferred Conditions

Extremophile Type	Conditions for Optimal Growth
Acidophiles	pH 3 or below
Alkaliphiles	pH 9 or above
Thermophiles	Temperature 60–80 °C (140–176 °F)
Hyperthermophiles	Temperature 80–122 °C (176–250 °F)
Psychrophiles	Temperature of -15 °C (5 °F) or lower
Halophiles	Salt concentration of at least 0.2 M
Osmophiles	High sugar concentration

Table 20.1

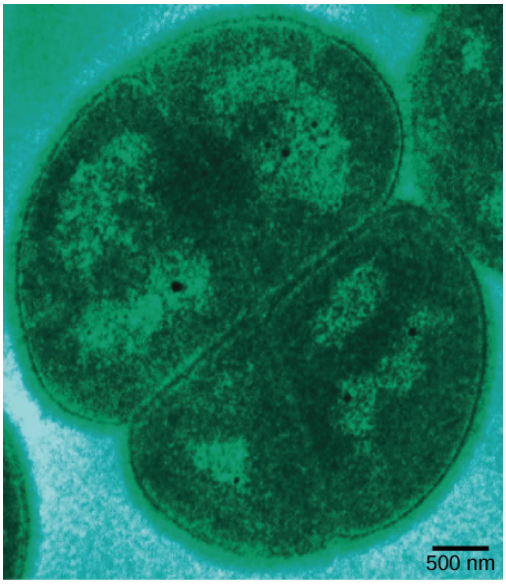


Figure 20.6 *Deinococcus radiodurans*, visualized in this false color transmission electron micrograph, is a prokaryote that can tolerate very high doses of ionizing radiation. It has developed DNA repair mechanisms that allow it to reconstruct its chromosome even if it has been broken into hundreds of pieces by radiation or heat. (credit: modification of work by Michael Daly; scale-bar data from Matt Russell)

The Ecology of Biofilms

Until a couple of decades ago, microbiologists used to think of prokaryotes as isolated entities living apart. This model, however, does not reflect the true ecology of prokaryotes, most of which prefer to live in communities where they can interact. A **biofilm** is a microbial community (**Figure 20.7**) held together in a gummy-textured matrix that consists primarily of polysaccharides secreted by the organisms, together with some proteins and nucleic acids. Biofilms grow attached to surfaces. Some of the best-studied biofilms are composed of prokaryotes, although fungal biofilms have also been described as well as some composed of a mixture of fungi and bacteria.

Biofilms are present almost everywhere: they can cause the clogging of pipes and readily colonize surfaces in industrial settings. In recent, large-scale outbreaks of bacterial contamination of food, biofilms have played a major role. They also colonize household surfaces, such as kitchen counters, cutting boards, sinks, and toilets, as well as places on the human body, such as the surfaces of our teeth.

Interactions among the organisms that populate a biofilm, together with their protective exopolysaccharidic (EPS) environment, make these communities more robust than free-living, or planktonic, prokaryotes. The sticky substance that holds bacteria together also excludes most antibiotics and disinfectants, making biofilm bacteria hardier than their planktonic counterparts. Overall, biofilms are very difficult to destroy because they are resistant to many common forms of sterilization.

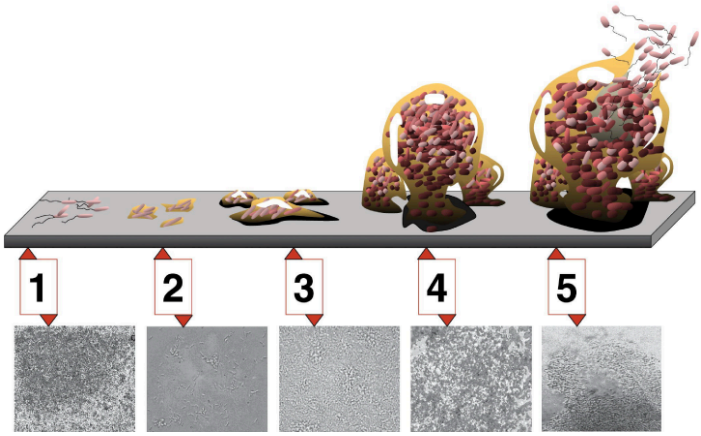
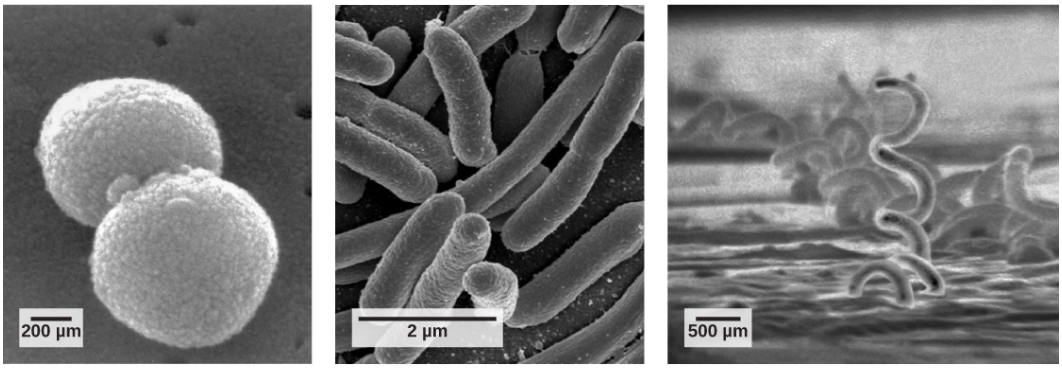


Figure 20.7 Five stages of biofilm development are shown. During stage 1, initial attachment, bacteria adhere to a solid surface via weak van der Waals interactions. During stage 2, irreversible attachment, hairlike appendages called pili permanently anchor the bacteria to the surface. During stage 3, maturation I, the biofilm grows through cell division and recruitment of other bacteria. An extracellular matrix composed primarily of polysaccharides holds the biofilm together. During stage 4, maturation II, the biofilm continues to grow and takes on a more complex shape. During stage 5, dispersal, the biofilm matrix is partly broken down, allowing some bacteria to escape and colonize another surface. Micrographs of a *Pseudomonas aeruginosa* biofilm in each of the stages of development are shown. (credit: D. Davis, Don Monroe, PLoS)

Structure of Prokaryotes

There are many differences between prokaryotic and eukaryotic cells. However, all cells have four common structures: the plasma membrane, which functions as a barrier for the cell and separates the cell from its environment; the cytoplasm, a jelly-like substance inside the cell; nucleic acids, the genetic material of the cell; and ribosomes, where protein synthesis takes place. Prokaryotes come in various shapes, but many fall into three categories: cocci (spherical), bacilli (rod-shaped), and spirilli (spiral-shaped) (**Figure 20.8**).



(a) (b) (c)

Figure 20.8 Prokaryotes fall into three basic categories based on their shape, visualized here using scanning electron microscopy: (a) cocci, or spherical (a pair is shown); (b) bacilli, or rod-shaped; and (c) spirilli, or spiral-shaped. (credit a: modification of work by Janice Haney Carr, Dr. Richard Facklam, CDC; credit c: modification of work by Dr. David Cox; scale-bar data from Matt Russell)

The Prokaryotic Cell

Recall that prokaryotes (**Figure 20.9**) are unicellular organisms that lack organelles or other internal membrane-bound structures. Therefore, they do not have a nucleus but instead generally have a single chromosome—a piece of circular, double-stranded DNA located in an area of the cell called the nucleoid. Most prokaryotes have a cell wall outside the plasma membrane.

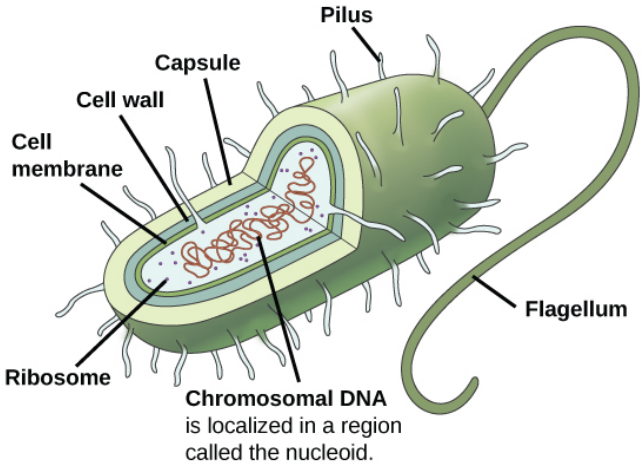


Figure 20.9 The features of a typical prokaryotic cell are shown.

Recall that prokaryotes are divided into two different domains, Bacteria and Archaea, which together with Eukarya, comprise the three domains of life (**Figure 20.10**).

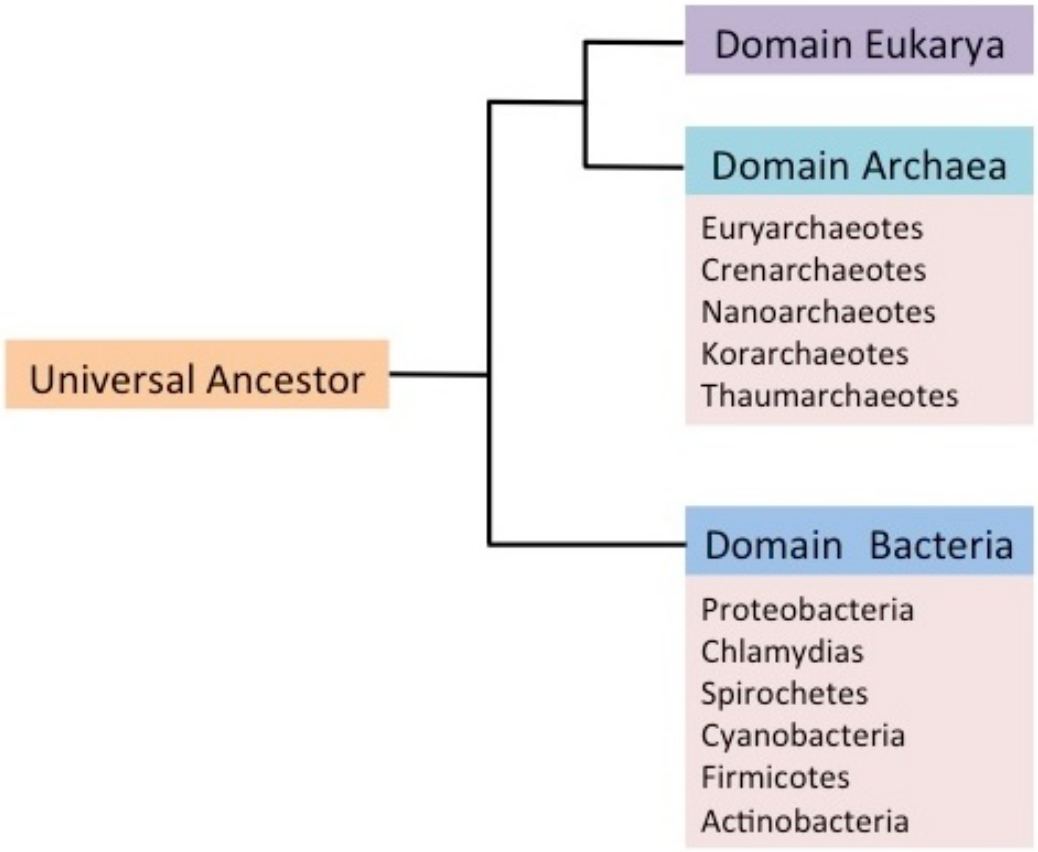


Figure 20.10 Bacteria and Archaea are both prokaryotes but differ enough to be placed in separate domains. An ancestor of modern Archaea is believed to have given rise to Eukarya, the third domain of life. Archaeal and bacterial phyla are shown; the evolutionary relationship between these phyla is still open to debate. Work by Robert A. Bear

The composition of the cell wall differs significantly between the domains Bacteria and Archaea. The composition of their cell walls also differs from the eukaryotic cell walls found in plants (cellulose) or fungi and insects (chitin). The cell wall functions as a protective layer, and it is responsible for the organism's shape. Some bacteria have an outer **capsule** outside the cell wall. Other structures are present in some prokaryotic species, but not in others (**m47338 (http://legacy.cnx.org/content/m47338/1.7/#tab-ch22-02-01)**). For example, the capsule found in some species enables the organism to attach to surfaces, protects it from dehydration and attack by phagocytic cells, and makes pathogens more resistant to our immune

responses. Some species also have flagella (singular, flagellum) used for locomotion, and **pili** (singular, pilus) used for attachment to surfaces. Plasmids, which consist of extra-chromosomal DNA, are also present in many species of bacteria and archaea.

Reproduction

Reproduction in prokaryotes is asexual and usually takes place by binary fission. Recall that the DNA of a prokaryote exists as a single, circular chromosome. Prokaryotes do not undergo mitosis. Rather the chromosome is replicated and the two resulting copies separate from one another, due to the growth of the cell. The prokaryote, now enlarged, is pinched inward at its equator and the two resulting cells, which are clones, separate. Binary fission does not provide an opportunity for genetic recombination or genetic diversity, but prokaryotes can share genes by three other mechanisms.

In **transformation**, the prokaryote takes in DNA found in its environment that is shed by other prokaryotes. If a nonpathogenic bacterium takes up DNA for a toxin gene from a pathogen and incorporates the new DNA into its own chromosome, it too may become pathogenic. In **transduction**, bacteriophages, the viruses that infect bacteria, sometimes also move short pieces of chromosomal DNA from one bacterium to another. Transduction results in a recombinant organism. Archaea are not affected by bacteriophages but instead have their own viruses that translocate genetic material from one individual to another. In **conjugation**, DNA is transferred from one prokaryote to another by means of a pilus, which brings the organisms into contact with one another. The DNA transferred can be in the form of a plasmid or as a hybrid, containing both plasmid and chromosomal DNA. These three processes of DNA exchange are shown in **m47338 (http://legacy.cnx.org/content/m47338/1.7/#fig-ch22_02_08)** .

Reproduction can be very rapid: a few minutes for some species. This short generation time coupled with mechanisms of genetic recombination and high rates of mutation result in the rapid evolution of prokaryotes, allowing them to respond to environmental changes (such as the introduction of an antibiotic) very quickly.

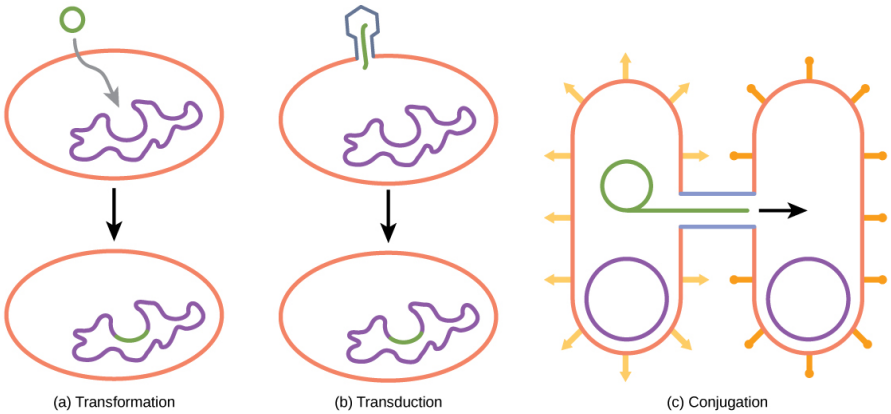


Figure 20.11 Besides binary fission, there are three other mechanisms by which prokaryotes can exchange DNA. In (a) transformation, the cell takes up prokaryotic DNA directly from the environment. The DNA may remain separate as plasmid DNA or be incorporated into the host genome. In (b) transduction, a bacteriophage injects DNA into the cell that contains a small fragment of DNA from a different prokaryote. In (c) conjugation, DNA is transferred from one cell to another via a pilus that connects the two cells.

evolution CONNECTION

The Evolution of Prokaryotes

How do scientists answer questions about the evolution of prokaryotes? Unlike with animals, artifacts in the fossil record of prokaryotes offer very little information. Fossils of ancient prokaryotes look like tiny bubbles in rock. Some scientists turn to genetics and to the principle of the molecular clock, which holds that the more recently two species have diverged, the more similar their genes (and thus proteins) will be. Conversely, species that diverged long ago will have more genes that are dissimilar.

Scientists at the NASA Astrobiology Institute and at the European Molecular Biology Laboratory collaborated to analyze the molecular evolution of 32 specific proteins common to 72 species of prokaryotes.[1] The model they derived from their data indicates that three important groups of bacteria—Actinobacteria, *Deinococcus*, and Cyanobacteria (which the authors call *Terrabacteria*)—were the first to colonize land. (Recall that *Deinococcus* is a genus of prokaryote—a bacterium—that is highly resistant to ionizing radiation.) Cyanobacteria are photosynthesizers, while Actinobacteria are a group of very common bacteria that include species important in decomposition of organic wastes.

The timelines of divergence suggest that bacteria (members of the domain Bacteria) diverged from common ancestral species between 2.5 and 3.2 billion years ago, whereas archaea diverged earlier: between 3.1 and 4.1 billion years ago. Eukarya later diverged off the Archaean line. The work further suggests that stromatolites that formed prior to the advent of cyanobacteria (about 2.6 billion years ago) photosynthesized in an anoxic environment and that because of the modifications of the Terrabacteria for land (resistance to drying and the possession of compounds that protect the organism from excess light), photosynthesis using oxygen may be closely linked to adaptations to survive on land.

Bacterial Diseases in Humans

Devastating pathogen-borne diseases and plagues, both viral and bacterial in nature, have affected humans since the beginning of human history. The true cause of these diseases was not understood at the time, and some people thought that diseases were a spiritual punishment. Over time, people came to realize that staying apart from afflicted persons, and disposing of the corpses and personal belongings of victims of illness, reduced their own chances of getting sick. **Table 20.2** below lists the disease and known associated bacterial cause.

Disease	Bacterium
Anthrax	*Bacillus anthracis*
Botulism	*Clostridium botulinum*
Chlamydia	*Chlamydia trachomatis*
Plague	*Yersinia pestis*
Tuberculosis	*Mycobacterium tuberculosis*
Peptic ulcers	*Helicobacter pylori*
Cholera	*Vibrio cholerae*
Lyme disease	*Borrelia burgdorferi*

Table 20.2 Human Disease and associated bacterium

Epidemiologists study how diseases affect a population. An **epidemic** is a disease that occurs in an unusually high number of individuals in a population at the same time. A **pandemic** is a widespread, usually worldwide, epidemic. An **endemic disease** is a disease that is constantly present, usually at low incidence, in a population.

1. Battistuzzi, FU, Feijao, A, and Hedges, SB. A genomic timescale of prokaryote evolution: Insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BioMed Central: Evolutionary Biology* 4 (2004): 44, doi:10.1186/1471-2148-4-44.

Long History of Bacterial Disease

There are records about infectious diseases as far back as 3000 B.C. A number of significant pandemics caused by bacteria have been documented over several hundred years. Some of the most memorable pandemics led to the decline of cities and nations.

In the 21st century, infectious diseases remain among the leading causes of death worldwide, despite advances made in medical research and treatments in recent decades. A disease spreads when the pathogen that causes it is passed from one person to another. For a pathogen to cause disease, it must be able to reproduce in the host's body and damage the host in some way.

The Plague of Athens

In 430 B.C., the Plague of Athens killed one-quarter of the Athenian troops that were fighting in the great Peloponnesian War and weakened Athens' dominance and power. The plague impacted people living in overcrowded Athens as well as troops aboard ships that had to return to Athens. The source of the plague may have been identified recently when researchers from the University of Athens were able to use DNA from teeth recovered from a mass grave. The scientists identified nucleotide sequences from a pathogenic bacterium, *Salmonella enterica* serovar Typhi (**Figure 20.12**), which causes typhoid fever.[2] This disease is commonly seen in overcrowded areas and has caused epidemics throughout recorded history.

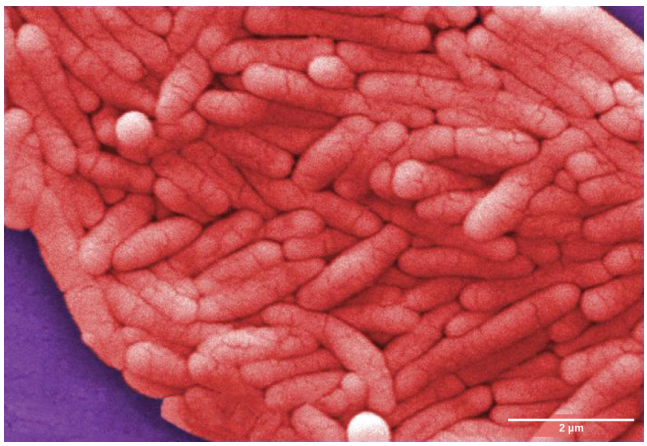


Figure 20.12 *Salmonella enterica* serovar Typhi, the causative agent of Typhoid fever, is a Gram-negative, rod-shaped gamma protobacterium. Typhoid fever, which is spread through feces, causes intestinal hemorrhage, high fever, delirium and dehydration. Today, between 16 and 33 million cases of this re-emerging disease occur annually, resulting in over 200,000 deaths. Carriers of the disease can be asymptomatic. In a famous case in the early 1900s, a cook named Mary Mallon unknowingly spread the disease to over fifty people, three of whom died. Other *Salmonella* serotypes cause food poisoning. (credit: modification of work by NCI, CDC)

Bubonic Plagues

From 541 to 750, an outbreak of what was likely a bubonic plague (the Plague of Justinian), eliminated one-quarter to one-half of the human population in the eastern Mediterranean region. The population in Europe dropped by 50 percent during this outbreak. Bubonic plague would strike Europe more than once.

One of the most devastating pandemics was the **Black Death** (1346 to 1361) that is believed to have been another outbreak of bubonic plague caused by the bacterium *Yersinia pestis*. It is thought to have been contracted initially in China and spread along the Silk Road, a network of land and sea trade routes, to the Mediterranean region and Europe, carried by rat fleas living on black rats that were always present on ships. The Black Death reduced the world's population from an estimated 450 million to about 350 to 375 million. Bubonic plague struck London hard again in the mid-1600s (**Figure 20.13**). In modern times, approximately 1,000 to 3,000 cases of plague arise globally each year. Although contracting bubonic plague before antibiotics meant almost certain death, the bacterium responds to several types of modern antibiotics, and mortality rates from plague are now very low.

2. Papagrigorakis MJ, Synodinos PN, and Yapijakis C. Ancient typhoid epidemic reveals possible ancestral strain of *Salmonella enterica* serovar Typhi. *Infect Genet Evol* 7 (2007): 126–7, Epub 2006 Jun.

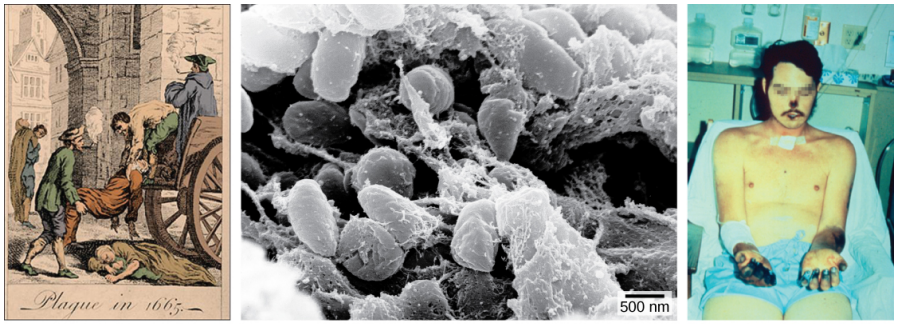


Figure 20.13 The (a) Great Plague of London killed an estimated 200,000 people, or about twenty percent of the city's population. The causative agent, the (b) bacterium *Yersinia pestis*, is a Gram-negative, rod-shaped bacterium from the class Gamma Proteobacteria. The disease is transmitted through the bite of an infected flea, which is infected by a rodent. Symptoms include swollen lymph nodes, fever, seizure, vomiting of blood, and (c) gangrene. (credit b: Rocky Mountain Laboratories, NIAID, NIH; scale-bar data from Matt Russell; credit c: Textbook of Military Medicine, Washington, D.C., U.S. Dept. of the Army, Office of the Surgeon General, Borden Institute)

Migration of Diseases to New Populations

Over the centuries, Europeans tended to develop genetic immunity to endemic infectious diseases, but when European conquerors reached the western hemisphere, they brought with them disease-causing bacteria and viruses, which triggered epidemics that completely devastated populations of Native Americans, who had no natural resistance to many European diseases. It has been estimated that up to 90 percent of Native Americans died from infectious diseases after the arrival of Europeans, making conquest of the New World a foregone conclusion.

Emerging and Re-emerging Diseases

The distribution of a particular disease is dynamic. Therefore, changes in the environment, the pathogen, or the host population can dramatically impact the spread of a disease. According to the World Health Organization (WHO) an **emerging disease** (**Figure 20.14**) is one that has appeared in a population for the first time, or that may have existed previously but is rapidly increasing in incidence or geographic range. This definition also includes re-emerging diseases that were previously under control. Approximately 75 percent of recently emerging infectious diseases affecting humans are zoonotic diseases, **zoonoses**, diseases that primarily infect animals and are transmitted to humans; some are of viral origin and some are of bacterial origin. Brucellosis is an example of a prokaryotic zoonosis that is re-emerging in some regions, and necrotizing fasciitis (commonly known as flesh-eating bacteria) has been increasing in virulence for the last 80 years for unknown reasons.

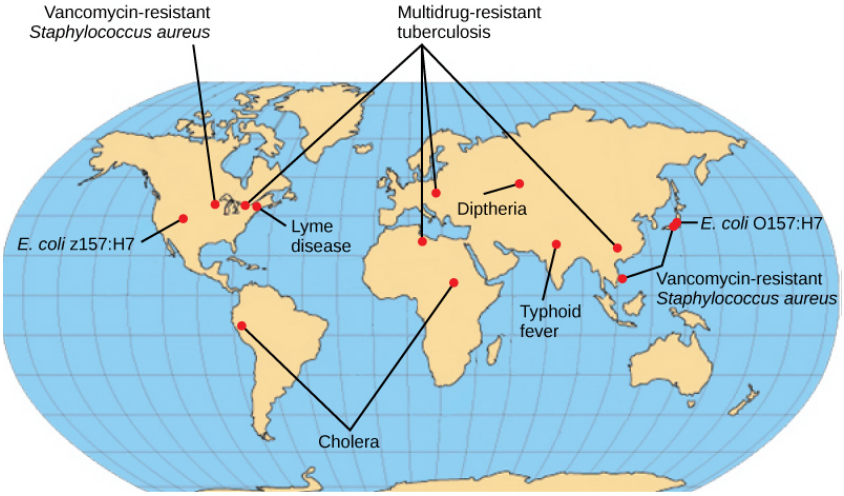
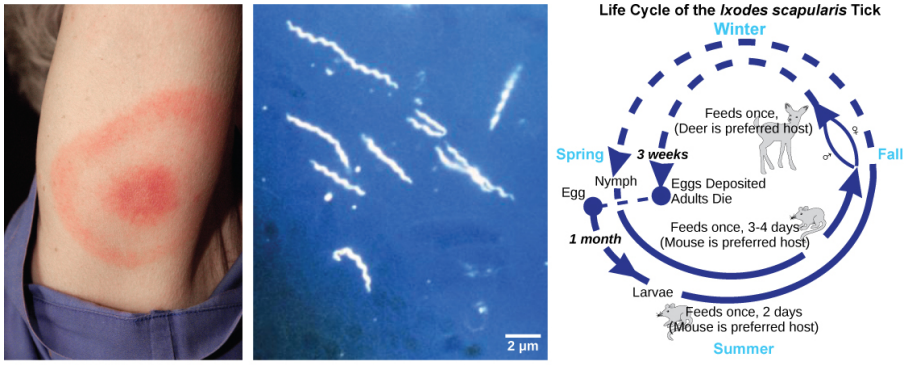


Figure 20.14 The map shows regions where bacterial diseases are emerging or reemerging. (credit: modification of work by NIH)

Some of the present emerging diseases are not actually new, but are diseases that were catastrophic in the past (**Figure 20.15**). They devastated populations and became dormant for a while, just to come back, sometimes more virulent than before, as was the case with bubonic plague. Other diseases, like tuberculosis, were never eradicated but were under control

in some regions of the world until coming back, mostly in urban centers with high concentrations of immunocompromised people. The WHO has identified certain diseases whose worldwide re-emergence should be monitored. Among these are two viral diseases (dengue fever and yellow fever), and three prokaryotic diseases (diphtheria, cholera, and bubonic plague). The war against infectious diseases has no foreseeable end.



(a) (b) (c)

Figure 20.15 Lyme disease often, but not always, results in (a) a characteristic bullseye rash. The disease is caused by a (b) Gram-negative spirochete bacterium of the genus *Borellia*. The bacteria (c) infect ticks, which in turns infect mice. Deer are the preferred secondary host, but the ticks also may feed on humans. Untreated, the disease causes chronic disorders in the nervous system, eyes, joints, and heart. The disease is named after Lyme, Connecticut, where an outbreak occurred in 1995 and has subsequently spread. The disease is not new, however. Genetic evidence suggests that Ötzi the Iceman, a 5,300-year-old mummy found in the Alps, was infected with *Borellia*. (credit a: James Gathany, CDC; credit b: CDC; scale-bar data from Matt Russell)

Biofilms and Disease

Recall that biofilms are microbial communities that are very difficult to destroy. They are responsible for diseases such as infections in patients with cystic fibrosis, Legionnaires' disease, and otitis media. They produce dental plaque and colonize catheters, prostheses, transcutaneous and orthopedic devices, contact lenses, and internal devices such as pacemakers. They also form in open wounds and burned tissue. In healthcare environments, biofilms grow on hemodialysis machines, mechanical ventilators, shunts, and other medical equipment. In fact, 65 percent of all infections acquired in the hospital (nosocomial infections) are attributed to biofilms. Biofilms are also related to diseases contracted from food because they colonize the surfaces of vegetable leaves and meat, as well as food-processing equipment that isn't adequately cleaned.

Biofilm infections develop gradually; sometimes, they do not cause symptoms immediately. They are rarely resolved by host defense mechanisms. Once an infection by a biofilm is established, it is very difficult to eradicate, because biofilms tend to be resistant to most of the methods used to control microbial growth, including antibiotics. Biofilms respond poorly or only temporarily to antibiotics; it has been said that they can resist up to 1,000 times the antibiotic concentrations used to kill the same bacteria when they are free-living or planktonic. An antibiotic dose that large would harm the patient; therefore, scientists are working on new ways to get rid of biofilms.

Antibiotics: Are We Facing a Crisis?

The word *antibiotic* comes from the Greek *anti* meaning "against" and *bios* meaning "life." An **antibiotic** is a chemical, produced either by microbes or synthetically, that is hostile to the growth of other organisms. Today's news and media often address concerns about an antibiotic crisis. Are the antibiotics that easily treated bacterial infections in the past becoming obsolete? Are there new "superbugs"—bacteria that have evolved to become more resistant to our arsenal of antibiotics? Is this the beginning of the end of antibiotics? All these questions challenge the healthcare community.

One of the main causes of resistant bacteria is the abuse of antibiotics. The imprudent and excessive use of antibiotics has resulted in the natural selection of resistant forms of bacteria. The antibiotic kills most of the infecting bacteria, and therefore only the resistant forms remain. These resistant forms reproduce, resulting in an increase in the proportion of resistant forms over non-resistant ones. Another major misuse of antibiotics is in patients with colds or the flu, for which antibiotics are useless. Another problem is the excessive use of antibiotics in livestock. The routine use of antibiotics in animal feed promotes bacterial resistance as well. In the United States, 70 percent of the antibiotics produced are fed to animals. These antibiotics are given to livestock in low doses, which maximize the probability of resistance developing, and these resistant bacteria are readily transferred to humans.

One of the Superbugs: MRSA

The imprudent use of antibiotics has paved the way for bacteria to expand populations of resistant forms. For example,

Staphylococcus aureus, often called “staph,” is a common bacterium that can live in the human body and is usually easily treated with antibiotics. A very dangerous strain, however, **methicillin-resistant *Staphylococcus aureus* (MRSA)** has made the news over the past few years (**Figure 20.16**). This strain is resistant to many commonly used antibiotics, including methicillin, amoxicillin, penicillin, and oxacillin. MRSA can cause infections of the skin, but it can also infect the bloodstream, lungs, urinary tract, or sites of injury. While MRSA infections are common among people in healthcare facilities, they have also appeared in healthy people who haven’t been hospitalized but who live or work in tight populations (like military personnel and prisoners). Researchers have expressed concern about the way this latter source of MRSA targets a much younger population than those residing in care facilities. *The Journal of the American Medical Association* reported that, among MRSA-afflicted persons in healthcare facilities, the average age is 68, whereas people with “community-associated MRSA” (**CA-MRSA**) have an average age of 23.[3]

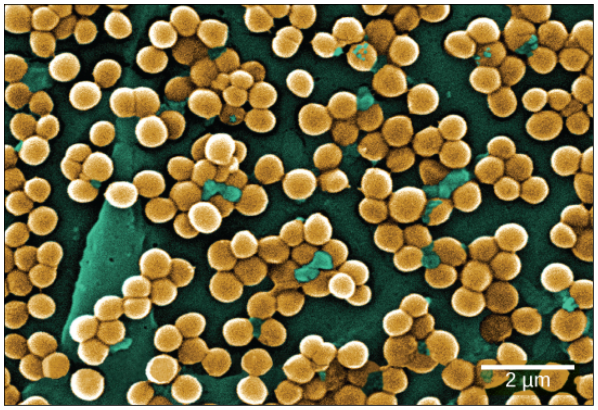


Figure 20.16 This scanning electron micrograph shows methicillin-resistant *Staphylococcus aureus* bacteria, commonly known as MRSA. *S. Aureus* is not always pathogenic, but can cause diseases such as food poisoning and skin and respiratory infections. (credit: modification of work by Janice Haney Carr; scale-bar data from Matt Russell)

In summary, the medical community is facing an antibiotic crisis. Some scientists believe that after years of being protected from bacterial infections by antibiotics, we may be returning to a time in which a simple bacterial infection could again devastate the human population. Researchers are developing new antibiotics, but it takes many years to of research and clinical trials, plus financial investments in the millions of dollars, to generate an effective and approved drug.

Foodborne Diseases

Prokaryotes are everywhere: They readily colonize the surface of any type of material, and food is not an exception. Most of the time, prokaryotes colonize food and food-processing equipment in the form of a biofilm. Outbreaks of bacterial infection related to food consumption are common. A **foodborne disease** (colloquially called “food poisoning”) is an illness resulting from the consumption of contaminated food, or the pathogenic bacteria, viruses, or other parasites that contaminate food. Although the United States has one of the safest food supplies in the world, the U.S. Centers for Disease Control and Prevention (CDC) has reported that “76 million people get sick, more than 300,000 are hospitalized, and 5,000 Americans die each year from foodborne illness.”

The characteristics of foodborne illnesses have changed over time. In the past, it was relatively common to hear about sporadic cases of **botulism**, the potentially fatal disease produced by a toxin from the anaerobic bacterium *Clostridium botulinum*. Some of the most common sources for this bacterium were non-acidic canned foods, homemade pickles, and processed meat and sausages. The can, jar, or package created a suitable anaerobic environment where *Clostridium* could grow. Proper sterilization and canning procedures have reduced the incidence of this disease.

While people may tend to think of foodborne illnesses as associated with animal-based foods, most cases are now linked to produce. There have been serious, produce-related outbreaks associated with raw spinach in the United States and with vegetable sprouts in Germany, and these types of outbreaks have become more common. The raw spinach outbreak in 2006 was produced by the bacterium *E. coli* serotype O157:H7. A **serotype** is a strain of bacteria that carries a set of similar antigens on its cell surface, and there are often many different serotypes of a bacterial species. Most *E. coli* are not particularly dangerous to humans, but serotype O157:H7 can cause bloody diarrhea and is potentially fatal.

All types of food can potentially be contaminated with bacteria. Recent outbreaks of *Salmonella* reported by the CDC occurred in foods as diverse as peanut butter, alfalfa sprouts, and eggs. A deadly outbreak in Germany in 2010 was caused by *E. coli* contamination of vegetable sprouts (**Figure 20.17**). The strain that caused the outbreak was found to be a new

3. Naimi, TS, LeDell, KH, Como-Sabetti, K, et al. Comparison of community- and health care-associated methicillin-resistant *Staphylococcus aureus* infection. *JAMA* 290 (2003): 2976–84, doi: 10.1001/jama.290.22.2976.

serotype not previously involved in other outbreaks, which indicates that *E. coli* is continuously evolving.

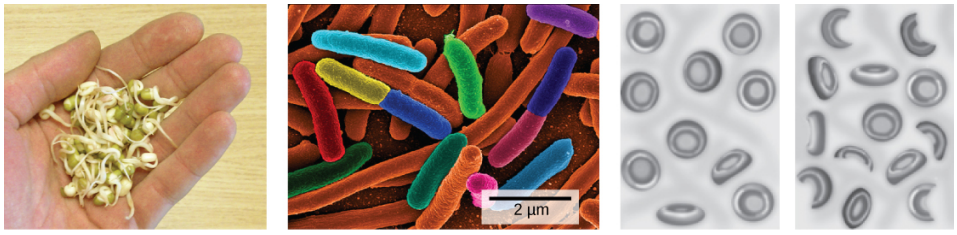


Figure 20.17 (a) Vegetable sprouts grown at an organic farm were the cause of an (b) *E. coli* outbreak that killed 32 people and sickened 3,800 in Germany in 2011. The strain responsible, *E. coli* O104:H4, produces Shiga toxin, a substance that inhibits protein synthesis in the host cell. The toxin (c) destroys red blood cells resulting in bloody diarrhea. Deformed red blood cells clog the capillaries of the kidney, which can lead to kidney failure, as happened to 845 patients in the 2011 outbreak. Kidney failure is usually reversible, but some patients experience kidney problems years later. (credit c: NIDDK, NIH)

career CONNECTION

Epidemiologist

Epidemiology is the study of the occurrence, distribution, and determinants of health and disease in a population. It is, therefore, part of public health. An epidemiologist studies the frequency and distribution of diseases within human populations and environments.

Epidemiologists collect data about a particular disease and track its spread to identify the original mode of transmission. They sometimes work in close collaboration with historians to try to understand the way a disease evolved geographically and over time, tracking the natural history of pathogens. They gather information from clinical records, patient interviews, surveillance, and any other available means. That information is used to develop strategies, such as vaccinations (**Figure 20.18**), and design public health policies to reduce the incidence of a disease or to prevent its spread. Epidemiologists also conduct rapid investigations in case of an outbreak to recommend immediate measures to control it.

An epidemiologist has a bachelor's degree, plus a master's degree in public health (MPH). Many epidemiologists are also physicians (and have an M.D.), or they have a Ph.D. in an associated field, such as biology or microbiology.

Figure 20.18 Vaccinations can slow the spread of communicable diseases. (credit: modification of work by Daniel Paquet)

Beneficial Bacteria

Not all prokaryotes are pathogenic. On the contrary, pathogens represent only a very small percentage of the diversity of the microbial world. In fact, our life would not be possible without prokaryotes. Just think about the role of prokaryotes in

biogeochemical cycles.

Cooperation between Bacteria and Eukaryotes: Nitrogen Fixation

Nitrogen is a very important element to living things, because it is part of nucleotides and amino acids that are the building blocks of nucleic acids and proteins, respectively. Nitrogen is usually the most limiting element in terrestrial ecosystems, with atmospheric nitrogen, N_2, providing the largest pool of available nitrogen. However, eukaryotes cannot use atmospheric, gaseous nitrogen to synthesize macromolecules. Fortunately, nitrogen can be "fixed," meaning it is converted into ammonia (NH_3) either biologically or abiotically. Abiotic nitrogen fixation occurs as a result of lightning or by industrial processes.

Biological nitrogen fixation (BNF) is exclusively carried out by prokaryotes: soil bacteria, cyanobacteria, and *Frankia* spp. (filamentous bacteria interacting with actinorhizal plants such as alder, bayberry, and sweet fern). After photosynthesis, BNF is the second most important biological process on Earth. The equation representing the process is as follows

$$N_2 + 16ATP + 8e^- + 8H^+ \rightarrow 2NH_3 + 16ADP + 16Pi + H_2$$

where Pi stands for inorganic phosphate. The total fixed nitrogen through BNF is about 100 to 180 million metric tons per year. Biological processes contribute 65 percent of the nitrogen used in agriculture.

Cyanobacteria are the most important nitrogen fixers in aquatic environments. In soil, members of the genus *Clostridium* are examples of free-living, nitrogen-fixing bacteria. Other bacteria live symbiotically with legume plants, providing the most important source of BNF. Symbionts may fix more nitrogen in soils than free-living organisms by a factor of 10. Soil bacteria, collectively called rhizobia, are able to symbiotically interact with legumes to form **nodules**, specialized structures where nitrogen fixation occurs (**Figure 20.19**). Nitrogenase, the enzyme that fixes nitrogen, is inactivated by oxygen, so the nodule provides an oxygen-free area for nitrogen fixation to take place. This process provides a natural and inexpensive plant fertilizer, as it reduces atmospheric nitrogen into ammonia, which is easily usable by plants. The use of legumes is an excellent alternative to chemical fertilization and is of special interest to sustainable agriculture, which seeks to minimize the use of chemicals and conserve natural resources. Through symbiotic nitrogen fixation, the plant benefits from using an endless source of nitrogen: the atmosphere. Bacteria benefit from using photosynthates (carbohydrates produced during photosynthesis) from the plant and having a protected niche. Additionally, the soil benefits from being naturally fertilized. Therefore, the use of rhizobia as biofertilizers is a sustainable practice.

Why are legumes so important? Some, like soybeans, are key sources of agricultural protein. Some of the most important grain legumes are soybean, peanuts, peas, chickpeas, and beans. Other legumes, such as alfalfa, are used to feed cattle.

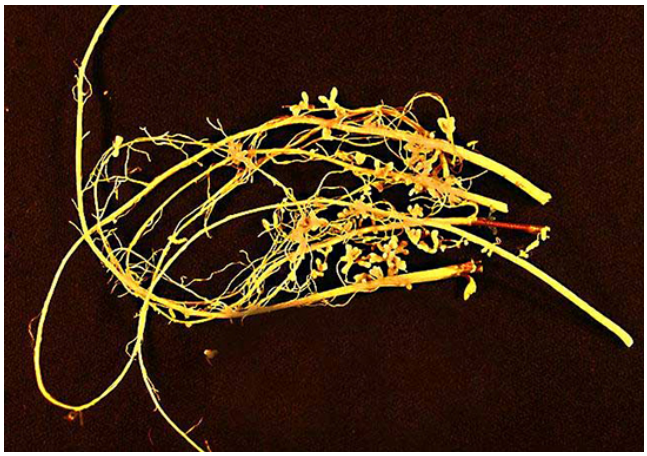


Figure 20.19 Soybean (*Glycine max*) is a legume that interacts symbiotically with the soil bacterium *Bradyrhizobium japonicum* to form specialized structures on the roots called nodules where nitrogen fixation occurs. (credit: USDA)

Early Biotechnology: Cheese, Bread, Wine, Beer, and Yogurt

According to the United Nations Convention on Biological Diversity, **biotechnology** is "any technological application that uses biological systems, living organisms, or derivatives thereof, to make or modify products or processes for specific use."[4] The concept of "specific use" involves some sort of commercial application. Genetic engineering, artificial selection, antibiotic production, and cell culture are current topics of study in biotechnology. However, humans have used prokaryotes before the term biotechnology was even coined. In addition, some of the goods and services are as simple as cheese, bread,

4. http://www.cbd.int/convention/articles/?a=cbd-02, United Nations Convention on Biological Diversity: Article 2: Use of Terms.

wine, beer, and yogurt, which employ both bacteria and other microbes, such as yeast, a fungus (Figure 20.20).

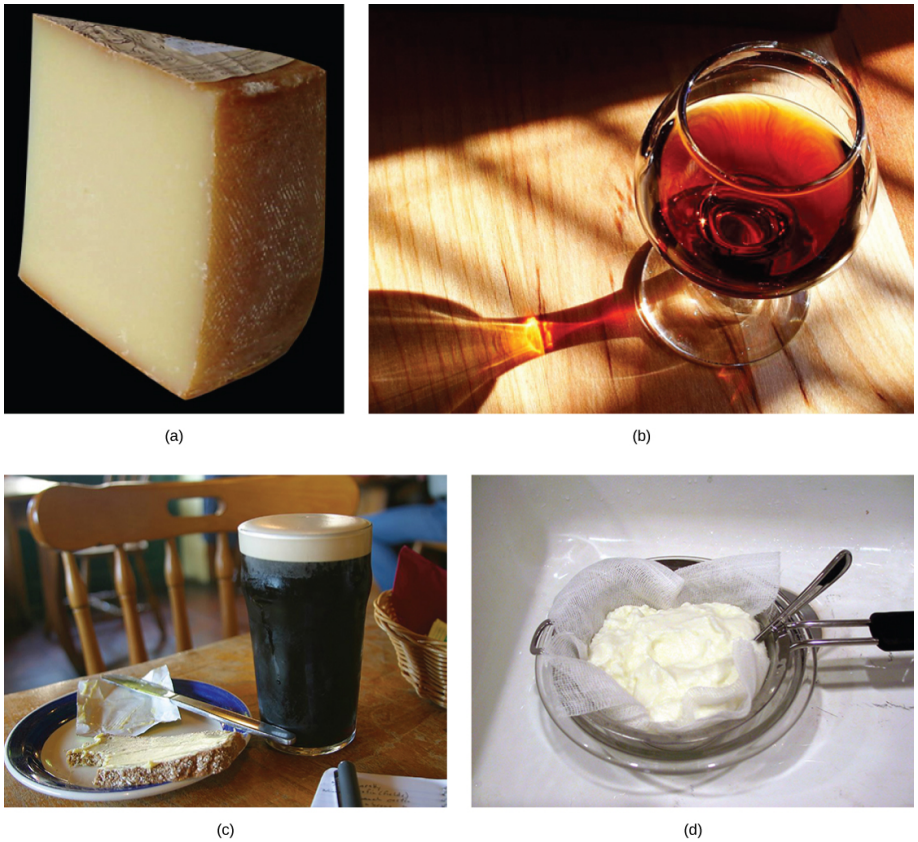


Figure 20.20 Some of the products derived from the use of prokaryotes in early biotechnology include (a) cheese, (b) wine, (c) beer and bread, and (d) yogurt. (credit bread: modification of work by F. Rodrigo/Wikimedia Commons; credit wine: modification of work by Jon Sullivan; credit beer and bread: modification of work by Kris Miller; credit yogurt: modification of work by Jon Sullivan)

Cheese production began around 4,000–7,000 years ago when humans began to breed animals and process their milk. Fermentation in this case preserves nutrients: Milk will spoil relatively quickly, but when processed as cheese, it is more stable. As for beer, the oldest records of brewing are about 6,000 years old and refer to the Sumerians. Evidence indicates that the Sumerians discovered fermentation by chance. Wine has been produced for about 4,500 years, and evidence suggests that cultured milk products, like yogurt, have existed for at least 4,000 years.

Using Prokaryotes to Clean up Our Planet: Bioremediation

Microbial **bioremediation** is the use of prokaryotes (or microbial metabolism) to remove pollutants. Bioremediation has been used to remove agricultural chemicals (pesticides, fertilizers) that leach from soil into groundwater and the subsurface. Certain toxic metals and oxides, such as selenium and arsenic compounds, can also be removed from water by bioremediation. The reduction of SeO_4^{-2} to SeO_3^{-2} and to Se^0 (metallic selenium) is a method used to remove selenium ions from water. Mercury is an example of a toxic metal that can be removed from an environment by bioremediation. As an active ingredient of some pesticides, mercury is used in industry and is also a by-product of certain processes, such as battery production. Methyl mercury is usually present in very low concentrations in natural environments, but it is highly toxic because it accumulates in living tissues. Several species of bacteria can carry out the biotransformation of toxic mercury into nontoxic forms. These bacteria, such as *Pseudomonas aeruginosa*, can convert Hg^{+2} into Hg^0, which is nontoxic to humans.

One of the most useful and interesting examples of the use of prokaryotes for bioremediation purposes is the cleanup of oil spills. The importance of prokaryotes to petroleum bioremediation has been demonstrated in several oil spills in recent years, such as the Exxon Valdez spill in Alaska (1989) (Figure 20.21), the Prestige oil spill in Spain (2002), the spill into the Mediterranean from a Lebanon power plant (2006), and more recently, the BP oil spill in the Gulf of Mexico (2010). To clean up these spills, bioremediation is promoted by the addition of inorganic nutrients that help bacteria to grow. Hydrocarbon-degrading bacteria feed on hydrocarbons in the oil droplet, breaking down the hydrocarbons. Some species,

such as *Alcanivorax borkumensis*, produce surfactants that solubilize the oil, whereas other bacteria degrade the oil into carbon dioxide. In the case of oil spills in the ocean, ongoing, natural bioremediation tends to occur, inasmuch as there are oil-consuming bacteria in the ocean prior to the spill. In addition to naturally occurring oil-degrading bacteria, humans select and engineer bacteria that possess the same capability with increased efficacy and spectrum of hydrocarbon compounds that can be processed. Under ideal conditions, it has been reported that up to 80 percent of the nonvolatile components in oil can be degraded within one year of the spill. Other oil fractions containing aromatic and highly branched hydrocarbon chains are more difficult to remove and remain in the environment for longer periods of time.

(a) (b)

Figure 20.21 (a) Cleaning up oil after the Valdez spill in Alaska, workers hosed oil from beaches and then used a floating boom to corral the oil, which was finally skimmed from the water surface. Some species of bacteria are able to solubilize and degrade the oil. (b) One of the most catastrophic consequences of oil spills is the damage to fauna. (credit a: modification of work by NOAA; credit b: modification of work by GOLUBENKOV, NGO: Saving Taman)

everyday CONNECTION

Microbes on the Human Body

The commensal bacteria that inhabit our skin and gastrointestinal tract do a host of good things for us. They protect us from pathogens, help us digest our food, and produce some of our vitamins and other nutrients. These activities have been known for a long time. More recently, scientists have gathered evidence that these bacteria may also help regulate our moods, influence our activity levels, and even help control weight by affecting our food choices and absorption patterns. The Human Microbiome Project has begun the process of cataloging our normal bacteria (and archaea) so we can better understand these functions.

A particularly fascinating example of our normal flora relates to our digestive systems. People who take high numbers of antibiotics tend to lose many of their normal gut bacteria, allowing a naturally antibiotic-resistant species called *Clostridium difficile* to overgrow and cause severe gastric problems, especially chronic diarrhea (**Figure 20.22**). Obviously, trying to treat this problem with antibiotics only makes it worse. However, it has been successfully treated by giving the patients fecal transplants from healthy donors to reestablish the normal intestinal microbial community. Clinical trials are underway to ensure the safety and effectiveness of this technique.

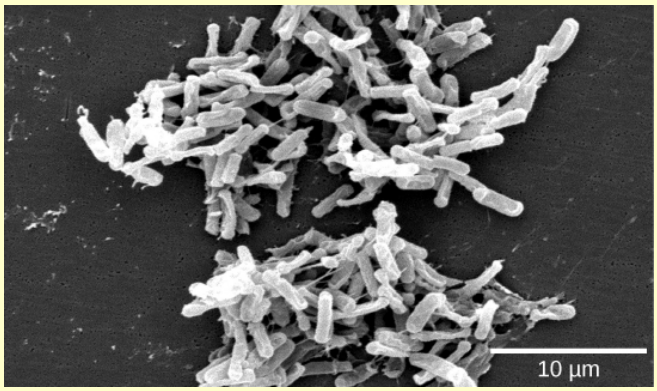


Figure 20.22 This scanning electron micrograph shows *Clostridium difficile*, a Gram-positive, rod-shaped bacterium that causes severe diarrhea. Infection commonly occurs after the normal gut fauna is eradicated by antibiotics. (credit: modification of work by CDC, HHS; scale-bar data from Matt Russell)

Scientists are also discovering that the absence of certain key microbes from our intestinal tract may set us up for a variety of problems. This seems to be particularly true regarding the appropriate functioning of the immune system. There are intriguing findings that suggest that the absence of these microbes is an important contributor to the development of allergies and some autoimmune disorders. Research is currently underway to test whether adding certain microbes to our internal ecosystem may help in the treatment of these problems as well as in treating some forms of autism.

20.3 | Kingdom Fungi

Introduction

> “Now when you cut a forest, an ancient forest in particular, you are not just removing a lot of big trees and a few birds fluttering around in the canopy. You are drastically imperiling a vast array of species within a few square miles of you. The number of these species may go to tens of thousands. ... Many of them are still

> unknown to science, and science has not yet discovered the key role undoubtedly played in the maintenance of that ecosystem, as in the case of fungi, microorganisms, and many of the insects."

Edward O. Wilson, "On Human Nature" (2000). In John H. Morgan, *Naturally Good*, (2005)

(a) (b) (c)

Figure 20.23 The (a) familiar mushroom is only one type of fungus. The brightly colored fruiting bodies of this (b) coral fungus are displayed. This (c) electron micrograph shows the spore-bearing structures of *Aspergillus*, a type of toxic fungi found mostly in soil and plants. (credit a: modification of work by Chris Wee; credit b: modification of work by Cory Zanker; credit c: modification of work by Janice Haney Carr, Robert Simmons, CDC; scale-bar data from Matt Russell)

The word *fungus* comes from the Latin word for mushroom. Indeed, the familiar mushrooms are fungi, but there are many other types of fungi as well (**Figure 20.23**). The kingdom Fungi includes an enormous variety of living organisms collectively referred to as Eumycota, or true fungi. While scientists have identified about 100,000 species of fungi, this is only a fraction of the over 5 million species likely present on Earth. Edible mushrooms, yeasts, black mold, and *Penicillium notatum* (the producer of the antibiotic penicillin) are all members of the kingdom Fungi, which belongs to the domain Eukarya. As eukaryotes, a typical fungal cell contains a true nucleus and many membrane-bound organelles.

Fungi were once considered plant-like organisms; however, DNA comparisons have shown that fungi are more closely related to animals than plants. Fungi are not capable of photosynthesis: They use complex organic compounds as sources of energy and carbon. Some fungal organisms multiply only asexually, whereas others undergo both asexual reproduction and sexual reproduction. Most fungi produce a large number of spores that are disseminated by the wind. Like bacteria, fungi play an essential role in ecosystems, because they are decomposers and participate in the cycling of nutrients by breaking down organic materials into simple molecules.

Fungi often interact with other organisms, forming mutually beneficial or mutualistic associations. Fungi also cause serious infections in plants and animals. For example, Dutch elm disease is a particularly devastating fungal infection that destroys many native species of elm (*Ulmus* spp.). The fungus infects the vascular system of the tree. It was accidentally introduced to North America in the 1900s and decimated elm trees across the continent. Dutch elm disease is caused by the fungus *Ophiostoma ulmi*. The elm bark beetle acts as a vector and transmits the disease from tree to tree. Many European and Asiatic elms are less susceptible than American elms.

In humans, fungal infections are generally considered challenging to treat because, unlike bacteria, they do not respond to traditional antibiotic therapy since they are also eukaryotes. These infections may prove deadly for individuals with a compromised immune system.

Fungi have many commercial applications. The food industry uses yeasts in baking, brewing, and wine making. Many industrial compounds are byproducts of fungal fermentation. Fungi are the source of many commercial enzymes and antibiotics.

Cell Structure and Function

Fungi are eukaryotes and as such have a complex cellular organization. As eukaryotes, fungal cells contain a membrane-bound nucleus. A few types of fungi have structures comparable to the plasmids (loops of DNA) seen in bacteria. Fungal cells also contain mitochondria and a complex system of internal membranes, including the endoplasmic reticulum and

Golgi apparatus.

Fungal cells do not have chloroplasts. Although the photosynthetic pigment chlorophyll is absent, many fungi display bright colors, ranging from red to green to black. The poisonous *Amanita muscaria* (fly agaric) is recognizable by its bright red cap with white patches (**Figure 20.24**). Pigments in fungi are associated with the cell wall and play a protective role against ultraviolet radiation. Some pigments are toxic.

Figure 20.24 The poisonous *Amanita muscaria* is native to the temperate and boreal regions of North America. (credit: Christine Majul)

Like plant cells, fungal cells are surrounded by a thick cell wall; however, the rigid layers contain the complex polysaccharides chitin and glucan. Cellulose, the main component of plant cell walls, is found rarely in fungi. Chitin, also found in the exoskeleton of insects, gives structural strength to the cell walls of fungi. The cell wall protects the cell from desiccation and predators. Similar to plants, some fungi contain a large central vacuole. The large central vacuole permits growth without spending many resources on regenerating the expensive cytoplasm. Fungi have plasma membranes similar to other eukaryotes, except that the structure is stabilized by ergosterol, a steroid molecule that functions like the cholesterol found in animal plasma membranes. Most members of the kingdom Fungi are nonmotile. Flagella are produced only by the gametes in the primitive division Chytridiomycota.

Growth

The vegetative body of a fungus is called a thallus and can be unicellular or multicellular. Some fungi are dimorphic because they can go from being unicellular to multicellular depending on environmental conditions. Unicellular fungi are generally referred to as **yeasts**.*Saccharomyces cerevisiae* (baker's yeast) and *Candida* species (the agents of thrush, a common fungal infection) are examples of unicellular fungi.

Most fungi are multicellular organisms. They display two distinct morphological stages: vegetative and reproductive. The vegetative stage is characterized by a tangle of slender thread-like structures called hyphae (singular, **hypha**), whereas the reproductive stage can be more conspicuous. A mass of hyphae is called a **mycelium** (**Figure 20.25**). It can grow on a surface, in soil or decaying material, in a liquid, or even in or on living tissue. Although individual hypha must be observed under a microscope, the mycelium of a fungus can be very large with some species truly being "the fungus humongous." The giant *Armillaria ostoyae* (honey mushroom) is considered the largest organism on Earth, spreading across over 2,000 acres of underground soil in eastern Oregon; it is estimated to be at least 2,400 years old.

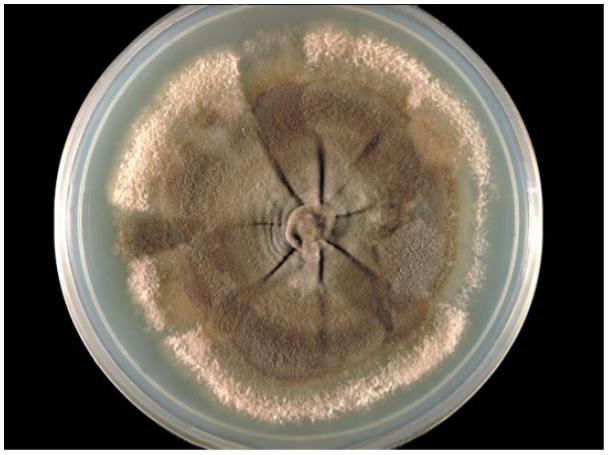


Figure 20.25 The mycelium of the fungus *Neotestudina rosati* can be pathogenic to humans. The fungus enters through a cut or scrape and develops into a mycetoma, a chronic subcutaneous infection. (credit: CDC)

Most fungal hyphae are divided into separate cells by end walls called septa (singular, septum. In most divisions (like plants, fungal phyla are called *divisions* by tradition) of fungi, tiny holes in the septa allow for the rapid flow of nutrients and small molecules from cell to cell along the hyphae. They are described as perforated septa. The hyphae in bread molds (which belong to the division Zygomycota) are not separated by septa. They are formed of large cells containing many nuclei, an arrangement described as coenocytic hyphae.

Fungi thrive in environments that are moist and slightly acidic, and can grow with or without light. They vary in their oxygen requirements. Most fungi are obligate aerobes, requiring oxygen to survive. Other species, such as the Chytridiomycota that reside in the rumen of cattle, are obligate anaerobes, meaning that they cannot grow and reproduce in an environment with oxygen. Yeasts are facultative: They grow best in the presence of oxygen but can use fermentation in the absence of oxygen. The alcohol produced from yeast fermentation is used in wine and beer production, and the carbon dioxide they produce carbonates beer and sparkling wine, and makes bread rise.

How Fungi Obtain Nutrition

Like animals, fungi are heterotrophs: They use complex organic compounds as a source of carbon rather than fixing carbon dioxide from the atmosphere, as some bacteria and most plants do. In addition, fungi do not fix nitrogen from the atmosphere. Like animals, they must obtain it from their diet. However, unlike most animals that ingest food and then digest it internally in specialized organs, fungi perform these steps in the reverse order. Digestion precedes ingestion. First, exoenzymes, enzymes that catalyze reactions on compounds outside of the cell, are transported out of the hyphae where they break down nutrients in the environment. Then, the smaller molecules produced by the external digestion are absorbed through the large surface areas of the mycelium. As with animal cells, the fungal storage polysaccharide is glycogen rather than starch, as found in plants.

Fungi are mostly **saprobes**, organisms that derive nutrients from decaying organic matter. They obtain their nutrients from dead or decomposing organic matter, mainly plant material. Fungal exoenzymes are able to break down insoluble polysaccharides, such as the cellulose and lignin of dead wood, into readily absorbable glucose molecules. Decomposers are important components of ecosystems, because they return nutrients locked in dead bodies to a form that is usable for other organisms. This role is discussed in more detail later. Because of their varied metabolic pathways, fungi fulfill an important ecological role and are being investigated as potential tools in bioremediation. For example, some species of fungi can be used to break down diesel oil and polycyclic aromatic hydrocarbons. Other species take up heavy metals such as cadmium and lead.

Some fungi are **parasitic** and use enzymes systems similar to the ones described above to break down plant cell walls and access plant resources within the cell. Other parasitic fungi can also penetrate the outside defenses of animals and acquire nourishment from the organism. These fungi are pathogenic and cause disease in the host see the pathogenic section for more information.

Reproduction

Fungi reproduce sexually and/or asexually. When Fungi reproduce sexually, spores are produced by meiosis, and these spores are referred to as meiospores. Fungi can also produce spores mitosis (asexual reproduction), and the spores produced are called mitospores. There are groups of fungi that reproduce only using mitospores are called the mitosporic fungi.

In both sexual and asexual reproduction, fungi produce spores that disperse from the parent organism by either floating

on the wind or hitching a ride on an animal. Fungal spores are smaller and lighter than plant seeds. The giant puffball mushroom bursts open and releases trillions of spores. The huge number of spores released increases the likelihood of landing in an environment that will support growth (**Figure 20.26**).

(a) (b)

Figure 20.26 The (a) giant puff ball mushroom releases (b) a cloud of spores when it reaches maturity. (credit a: modification of work by Roger Griffith; credit b: modification of work by Pearson Scott Foresman, donated to the Wikimedia Foundation)

Asexual Reproduction

Fungi reproduce asexually by fragmentation of the mycelium. Fragments of hyphae can grow new colonies which are identical. The fragmentation of mycelium maintain clonal populations adapted to a specific niche, and allows for more rapid dispersal than sexual reproduction.

The most common mode of asexual reproduction is through the formation of asexual spores or mitospores, which are produced by one parent only (through mitosis) and are genetically identical to that parent (**Figure 20.27**). Spores allow fungi to expand their distribution and colonize new environments. They may be released from the parent thallus either outside or within a specialized reproductive structure.

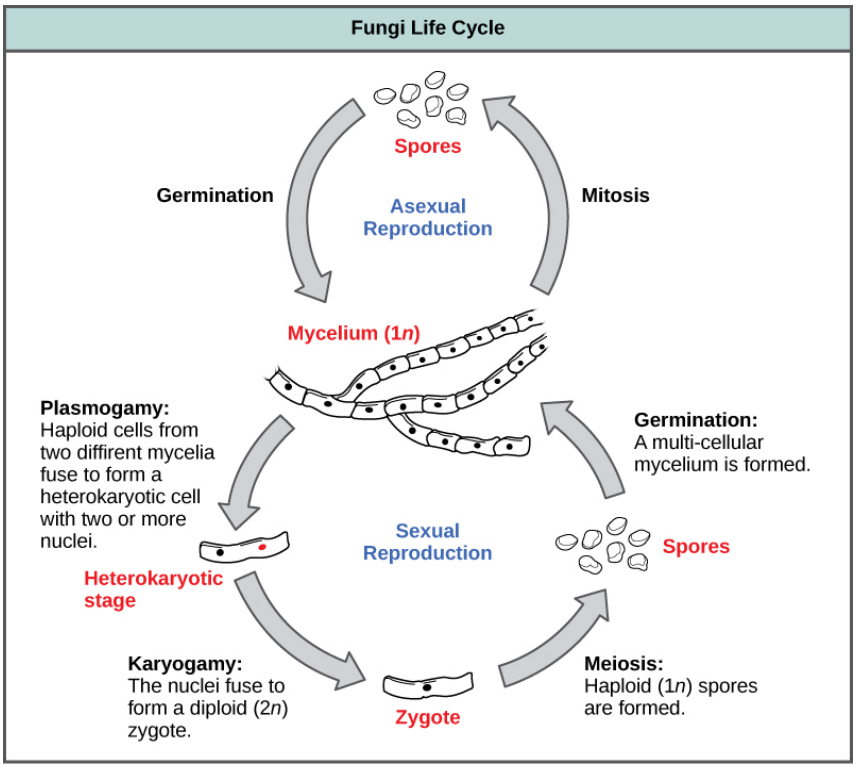


Figure 20.27 Fungi may have both asexual and sexual stages of reproduction.

There are many types of mitospores. Conidiospores are unicellular or multicellular spores that are released directly from

the tip or side of the hypha. Other mitospores originate in the fragmentation of a hypha to form single cells that are released as spores; some of these have a thick wall surrounding the fragment. Yet others bud off the vegetative parent cell. Sporangiospores are produced in a sporangium (**Figure 20.28**).

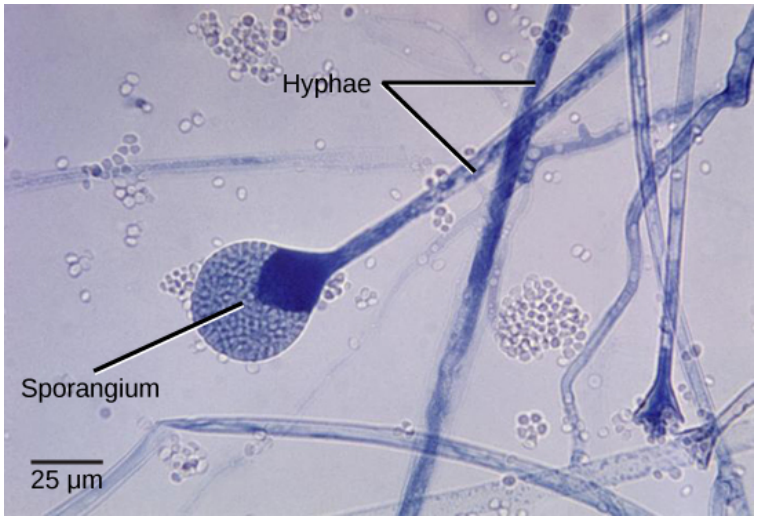


Figure 20.28 This bright field light micrograph shows the release of spores from a sporangium at the end of a hypha called a sporangiophore. The organism is a *Mucor* sp. fungus, a mold often found indoors. (credit: modification of work by Dr. Lucille Georg, CDC; scale-bar data from Matt Russell)

Sexual Reproduction

Sexual reproduction introduces genetic variation into a population of fungi. In fungi, sexual reproduction often occurs in response to adverse environmental conditions. During sexual reproduction, two mating types are produced. When both mating types are present in the same mycelium, it is called homothallic, or self-fertile. Heterothallic mycelia require two different, but compatible, mycelia to reproduce sexually.

Although there are many variations in fungal sexual reproduction, all include the following three stages (**Figure 20.27**). First, during cytoplasmic fusion (plasmogamy), two haploid cells fuse, leading to a dikaryotic stage where two haploid nuclei coexist in a single cell. During karyogamy (“nuclear marriage”), the haploid nuclei fuse to form a diploid zygote nucleus. For some groups of fungi, karyogamy and plasmogamy can be separated for a long periods of time and some fungi maintain dikaryotic vegetative mycelium. Finally, meiosis takes place in the gametangia (singular, gametangium) organs, in which gametes of different mating types are generated. At this stage, spores are disseminated into the environment.

Fungal Diversity

The kingdom Fungi contains five major phyla that were established according to their mode of sexual reproduction or using molecular data. Polyphyletic, unrelated fungi that reproduce without a sexual cycle, are placed for convenience in a sixth group called a “form phylum”. Not all mycologists agree with this scheme. Rapid advances in molecular biology and the sequencing of rRNA genes continue to show new and different relationships between the various categories of fungi (**Figure 20.29**). The five true phyla of fungi are the Chytridiomycota (Chytrids), the Zygomycota (conjugated fungi), the Ascomycota (sac fungi), the Basidiomycota (club fungi) and the recently described Phylum Glomeromycota. The Deuteromycota is an informal group of unrelated fungi that all share a common character – they use strictly asexual reproduction or are called mitosporic.

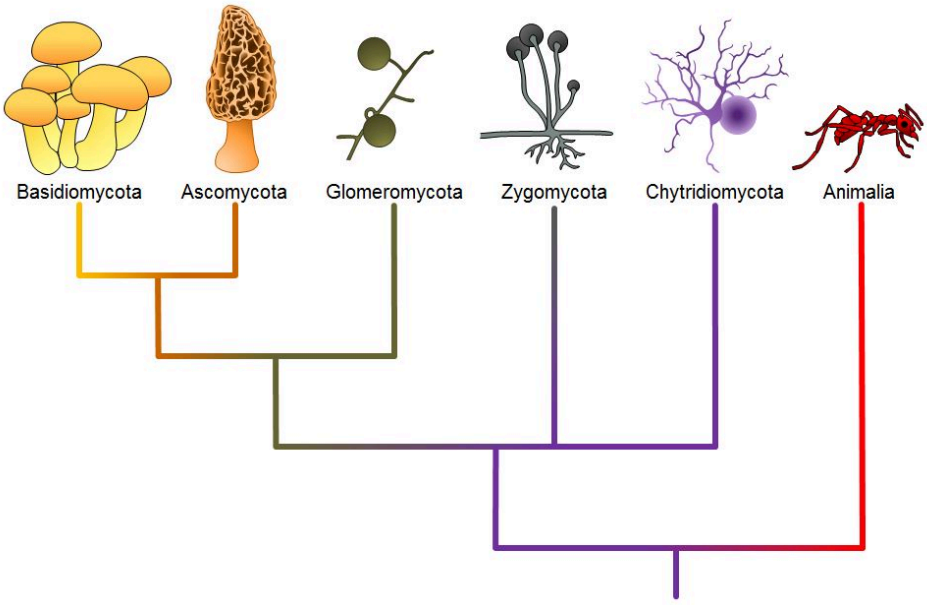


Figure 20.29 A phylogenetic tree of five groups of fungi Chytridiomycota (Chytrids), the Zygomycota (conjugated fungi), the Ascomycota (sac fungi), the Basidiomycota (club fungi) and the recently described Phylum Glomeromycota. Work by Eva Horne.

Note: "-mycota" is used to designate a phylum while "-mycetes" formally denotes a class or is used informally to refer to all members of the phylum.

Chytridiomycota: The Chytrids

The only class in the Phylum Chytridiomycota is the **Chytridiomycetes**. The chytrids are the simplest and most primitive Eumycota, or true fungi. The evolutionary record shows that the first recognizable chytrids appeared during the late pre-Cambrian period, more than 500 million years ago. Like all fungi, chytrids have chitin in their cell walls, but one group of chytrids has both cellulose and chitin in the cell wall. Most chytrids are unicellular; a few form multicellular organisms and hyphae, which have no septa between cells (coenocytic). They produce gametes and diploid zoospores that swim with the help of a single flagellum.

The ecological habitat and cell structure of chytrids have much in common with protists. Chytrids usually live in aquatic environments, although some species live on land. Some species thrive as parasites on plants, insects, or amphibians (**Figure 20.30**), while others are saprobes. The chytrid species *Allomyces* is well characterized as an experimental organism. Its reproductive cycle includes both asexual and sexual phases. *Allomyces* produces diploid or haploid flagellated zoospores in a sporangium.

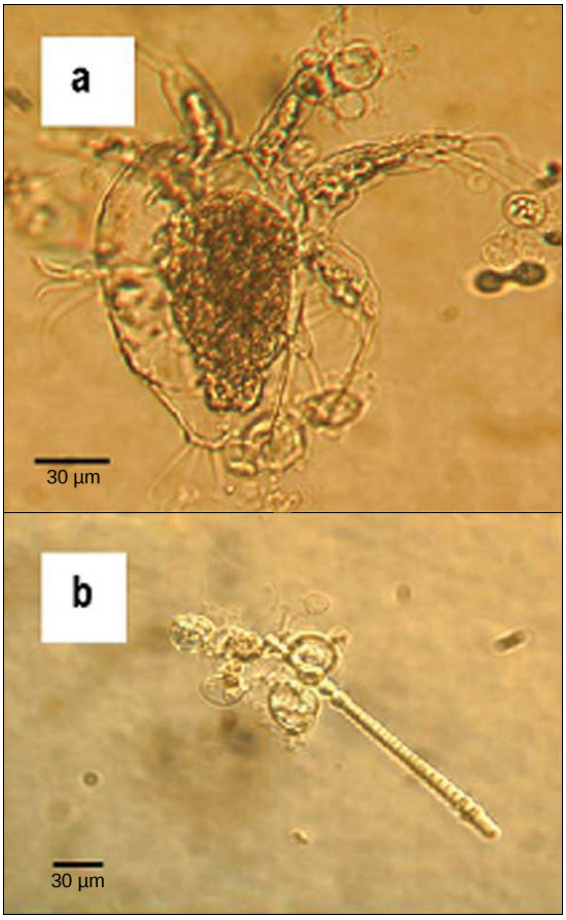


Figure 20.30 The chytrid *Batrachochytrium dendrobatidis* is seen in these light micrographs as transparent spheres growing on (a) a freshwater arthropod and (b) algae. This chytrid causes skin diseases in many species of amphibians, resulting in species decline and extinction. (credit: modification of work by Johnson ML, Speare R., CDC)

Zygomycota: The Conjugated Fungi

The zygomycetes are a relatively small group of fungi belonging to the Phylum **Zygomycota**. They include the familiar bread mold, *Rhizopus stolonifer*, which rapidly propagates on the surfaces of breads, fruits, and vegetables. Most species are saprobes, living off decaying organic material; a few are parasites, particularly of insects. Zygomycetes play a considerable commercial role. The metabolic products of other species of *Rhizopus* are intermediates in the synthesis of semi-synthetic steroid hormones.

Zygomycetes have a thallus of coenocytic hyphae in which the nuclei are haploid when the organism is in the vegetative stage. The fungi usually reproduce asexually by producing sporangiospores (**Figure 20.31**). The black tips of bread mold are the swollen sporangia packed with black spores (**Figure 20.32**). When spores land on a suitable substrate, they germinate and produce a new mycelium. Sexual reproduction starts when conditions become unfavorable. Two opposing mating strains (type + and type –) must be in close proximity for gametangia from the hyphae to be produced and fuse, leading to karyogamy. The developing diploid zygospores have thick coats that protect them from desiccation and other hazards. They may remain dormant until environmental conditions are favorable. When the zygospore germinates, it undergoes meiosis and produces haploid spores, which will, in turn, grow into a new organism. This form of sexual reproduction in fungi is called conjugation (although it differs markedly from conjugation in bacteria and protists), giving rise to the name “conjugated fungi”.

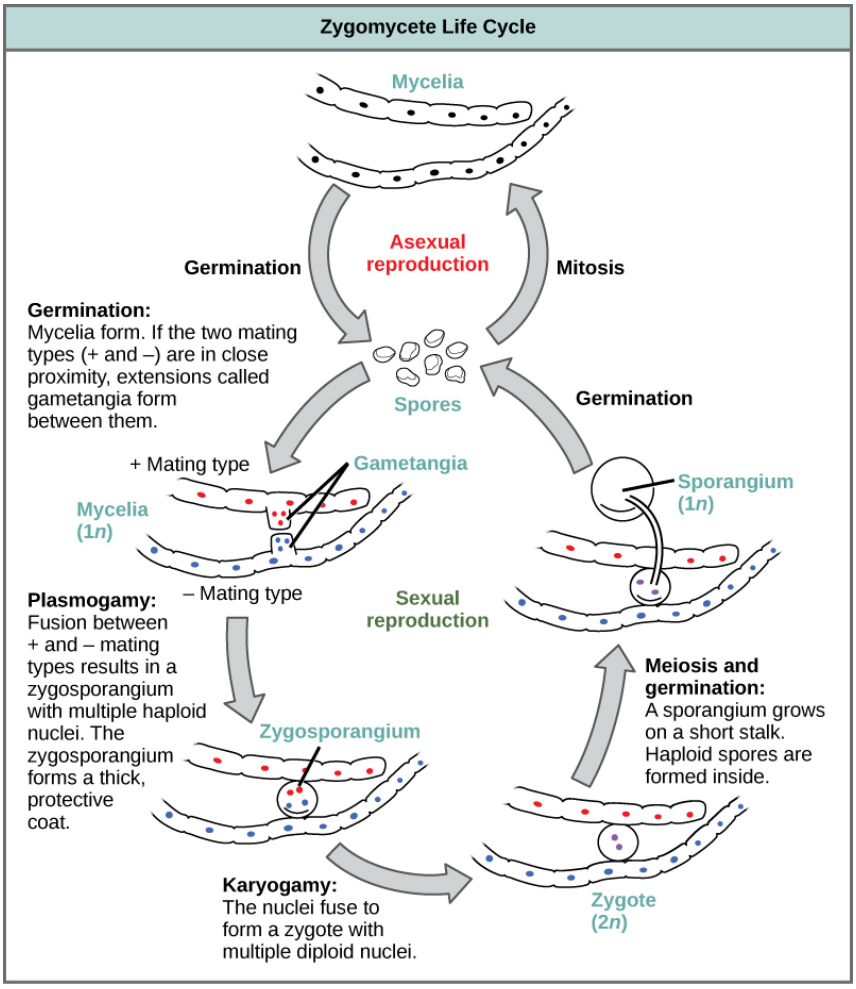
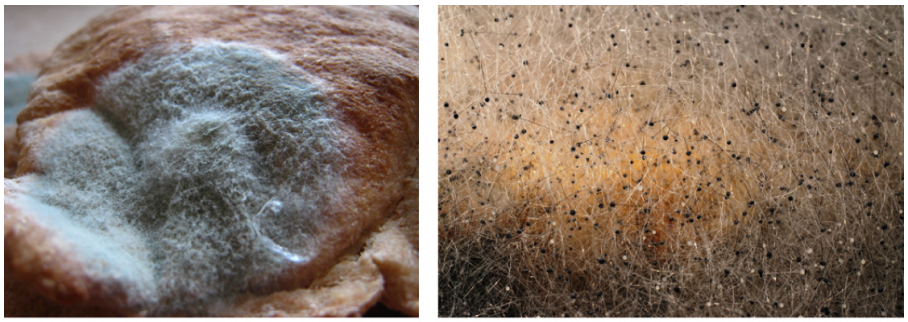


Figure 20.31 Zygomycetes have asexual and asexual life cycles. In the sexual life cycle, plus and minus mating types conjugate to form a zygosporangium.



(a) (b)

Figure 20.32 Sporangia grow at the end of stalks, which appear as (a) white fuzz seen on this bread mold, *Rhizopus stolonifer*. The (b) tips of bread mold are the spore-containing sporangia. (credit b: modification of work by "polandeze"/Flickr)

Ascomycota: The Sac Fungi

The majority of known fungi belong to the Phylum **Ascomycota**, which is characterized by the formation of an **ascus** (plural, asci), a sac-like structure that contains haploid ascospores. Many ascomycetes are of commercial importance. Some play a beneficial role, such as the yeasts used in baking, brewing, and wine fermentation, plus truffles and morels, which are held as gourmet delicacies. *Aspergillus oryzae* is used in the fermentation of rice to produce sake. Other ascomycetes parasitize plants and animals, including humans. For example, fungal pneumonia poses a significant threat to AIDS patients who have a compromised immune system. Ascomycetes not only infest and destroy crops directly; they also produce poisonous secondary metabolites that make crops unfit for consumption. Filamentous ascomycetes produce

hyphae divided by perforated septa, allowing streaming of cytoplasm from one cell to the other. Conidia and asci, which are used respectively for asexual and sexual reproductions, are usually separated from the vegetative hyphae by blocked (non-perforated) septa.

Asexual reproduction is frequent and involves the production of conidiophores that release haploid conidiospores (**Figure 20.33**). Sexual reproduction starts with the development of special hyphae from either one of two types of mating strains (**Figure 20.33**). The “male” strain produces an antheridium and the “female” strain develops an ascogonium. Next, cytoplasm of the antheridium and the ascogonium combine in plasmogamy without nuclear fusion. Special ascogenous hyphae arise, in which pairs of nuclei migrate: one from the “male” strain and one from the “female” strain. In each ascus, two or more haploid ascospores fuse their nuclei in karyogamy. During sexual reproduction, thousands of asci fill a fruiting body called the ascocarp. The diploid nucleus gives rise to haploid nuclei by meiosis. These haploid nuclei undergo mitosis and cell divisions to form haploid ascospores. The ascospores are then released, germinate, and form hyphae that are disseminated in the environment and start new mycelia (**Figure 20.34**).

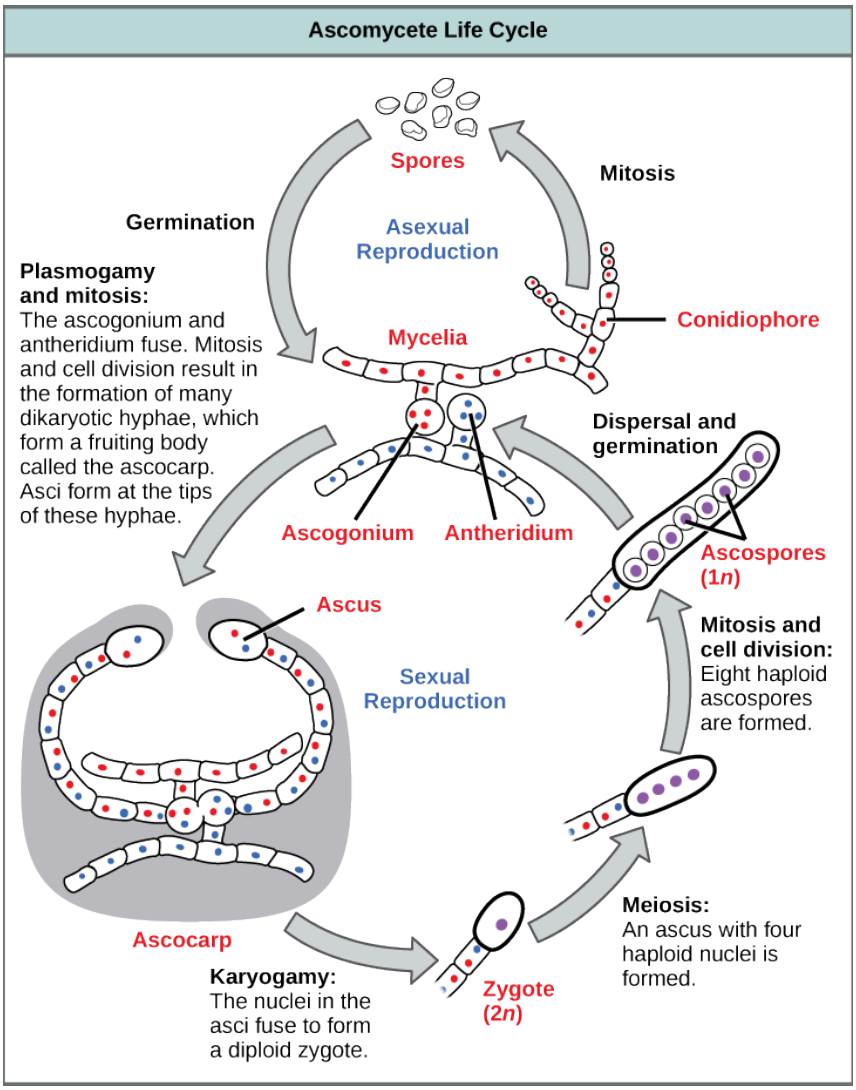


Figure 20.33 The lifecycle of an ascomycete is characterized by the production of asci during the sexual phase. The haploid phase is the predominant phase of the life cycle.

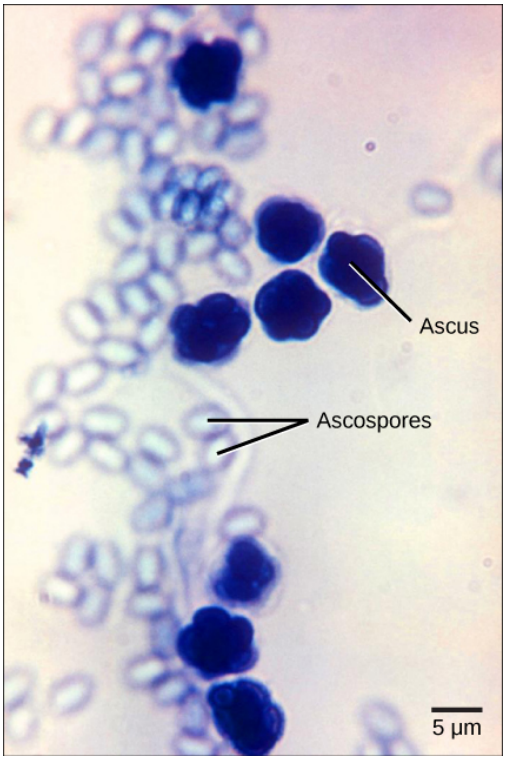


Figure 20.34 The bright field light micrograph shows ascospores being released from asci in the fungus *Talaromyces flavus* var. *flavus*. (credit: modification of work by Dr. Lucille Georg, CDC; scale-bar data from Matt Russell)

Basidiomycota: The Club Fungi

The fungi in the Phylum **Basidiomycota** are easily recognizable under a light microscope by their club-shaped fruiting bodies called **basidia** (singular, **basidium**), which are the swollen terminal cell of a hypha. This reproductive organ is the basidiocarp, that contains basidia, and the familiar mushroom, commonly seen in fields after rain, on the supermarket shelves, or growing on your lawn (**Figure 20.35**). These mushroom-producing basidiomyces are sometimes referred to as "gill fungi" because of the presence of gill-like structures on the underside of the cap. The "gills" are actually compacted hyphae on which the basidia are borne. This group also includes shelf fungus, which cling to the bark of trees like small shelves. In addition, the basidiomycota includes smuts and rusts, which are important plant pathogens; toadstools, and shelf fungi stacked on tree trunks. Most edible fungi belong to the Phylum Basidiomycota; however, some basidiomycetes produce deadly toxins. For example, *Cryptococcus neoformans* causes severe respiratory illness.

Figure 20.35 The fruiting bodies of a basidiomycete form a ring in a meadow, commonly called "fairy ring." The best-known fairy ring fungus has the scientific name *Marasmius oreades*. The body of this fungus, its mycelium, is underground and grows outward in a circle. As it grows, the mycelium depletes the soil of nitrogen, causing the mycelia to grow away from the center and leading to the "fairy ring" of fruiting bodies where there is adequate soil nitrogen. (Credit: "Cropcircles"/Wikipedia Commons)]

The lifecycle of basidiomycetes includes alternation of generations (Figure 20.36). Spores are generally produced through sexual reproduction, rather than asexual reproduction. The club-shaped basidium carries spores called basidiospores. In the basidium, nuclei of two different mating strains fuse (karyogamy), giving rise to a diploid zygote that then undergoes meiosis. The haploid nuclei migrate into basidiospores, which germinate and generate monokaryotic hyphae. The mycelium that results is called a primary mycelium. Mycelia of different mating strains can combine and produce a secondary mycelium that contains haploid nuclei of two different mating strains. This is the dikaryotic stage of the basidiomyces lifecyle and and it is the dominant stage. Eventually, the secondary mycelium generates a **basidiocarp**, which is a fruiting body that protrudes from the ground—this is what we think of as a mushroom. The basidiocarp bears the developing basidia on the gills under its cap.

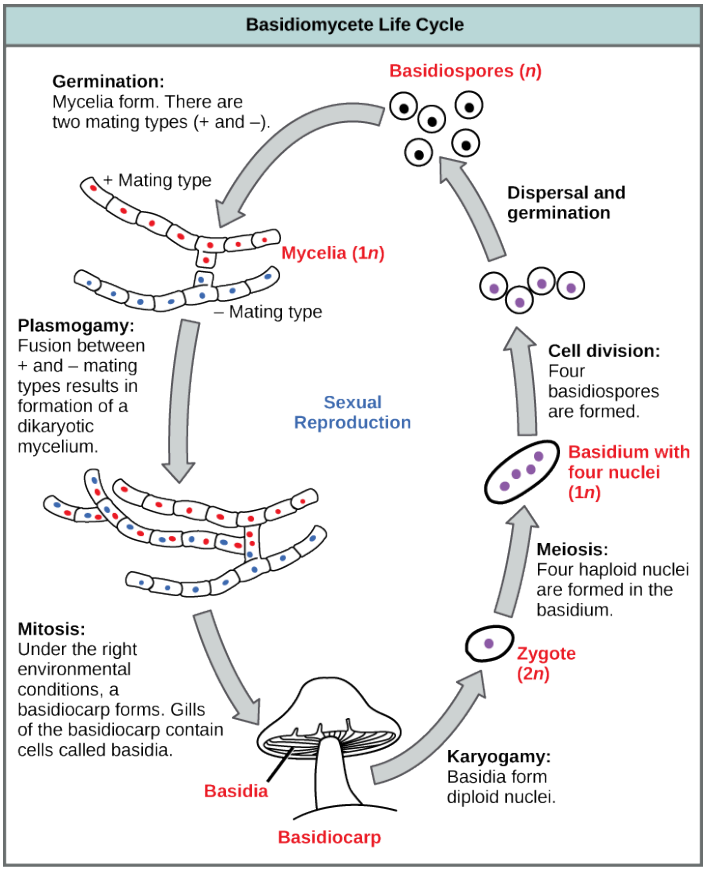


Figure 20.36 The lifecycle of a basidiomycete alternates generation with a prolonged stage in which two nuclei (dikaryon) are present in the hyphae.

Deuteromycota:

Those fungi that do not display a sexual phase and only reproduce by mitospores are classified in the form phylum **Deuteromycota**. Deuteromycota is a polyphyletic group where many species are more closely related to organisms in other phyla than to each other; hence it cannot be called a true phylum and must, instead, be given the name form phylum. Since they do not possess the sexual structures that are used to classify other fungi, they are less well described in comparison to other divisions. Most members live on land, with a few aquatic exceptions. They form visible mycelia with a fuzzy appearance and are commonly known as **mold**. Molecular analysis shows that the closest group to the deuteromycetes is the ascomycetes. In fact, some species, such as *Aspergillus*, which were once classified as imperfect fungi, are now classified as ascomycetes.

Reproduction of Deuteromycota is strictly asexual and occurs mostly by production of asexual conidiospores (Figure 20.37). Some hyphae may recombine and form heterokaryotic hyphae. Genetic recombination is known to take place between the different nuclei.

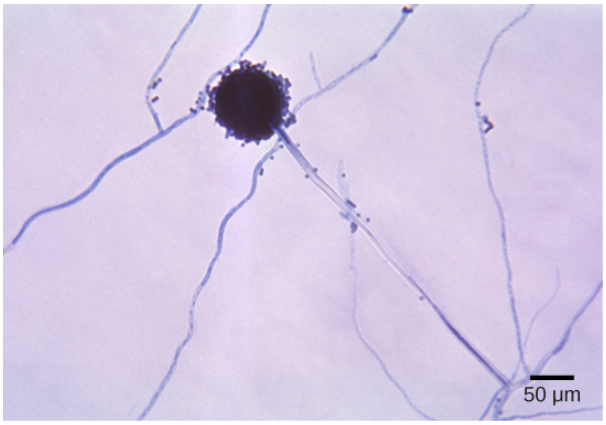


Figure 20.37 *Aspergillus niger* is an imperfect fungus commonly found as a food contaminant. The spherical structure in this light micrograph is a conidiophore. (credit: modification of work by Dr. Lucille Georg, CDC; scale-bar data from Matt Russell)

The Deuteromycetes have a large impact on everyday human life. The food industry relies on them for ripening some cheeses. The blue veins in Roquefort cheese and the white crust on Camembert are the result of fungal growth. The antibiotic penicillin was originally discovered on an overgrown Petri plate, on which a colony of *Penicillium* fungi killed the bacterial growth surrounding it. Many imperfect fungi cause serious diseases, either directly as parasites (which infect both plants and humans), or as producers of potent toxic compounds, as seen in the aflatoxins released by fungi of the genus *Aspergillus*.

Glomeromycota: The mitosporic fungi

The **Glomeromycota** is a newly established phylum which comprises about 230 species that all live in close association with the roots of trees. Fossil records indicate that trees and their root symbionts share a long evolutionary history. It appears that all members of this family form **arbuscular mycorrhizae**: the hyphae interact with the root cells forming a mutually beneficial association where the plants supply the carbon source and energy in the form of carbohydrates to the fungus, and the fungus supplies essential minerals from the soil to the plant.

The glomeromycetes do not reproduce sexually and do not survive without the presence of plant roots. Although they have coenocytic hyphae like the zygomycetes, they do not form zygospores. DNA analysis shows that all glomeromycetes probably descended from a common ancestor, making them a monophyletic lineage.

Pathogenic Fungi

Many fungi have negative impacts on other species, including humans and the organisms they depend on for food. Fungi may be parasites, pathogens, and, in a very few cases, predators.

Plant Parasites and Pathogens

The production of enough good-quality crops is essential to our existence. Plant diseases have ruined crops, bringing widespread famine. Most plant pathogens are fungi that cause tissue decay and eventual death of the host (**Figure 20.38**). In addition to destroying plant tissue directly, some plant pathogens spoil crops by producing potent toxins. Fungi are also responsible for food spoilage and the rotting of stored crops. For example, the fungus *Claviceps purpurea* causes ergot, a disease of cereal crops (especially of rye). Although the fungus reduces the yield of cereals, the effects of the ergot's alkaloid toxins on humans and animals are of much greater significance: In animals, the disease is referred to as ergotism. The most common signs and symptoms are convulsions, hallucination, gangrene, and loss of milk in cattle. The active ingredient of ergot is lysergic acid, which is a precursor of the drug LSD. Smuts, rusts, and powdery or downy mildew are other examples of common fungal pathogens that affect crops.

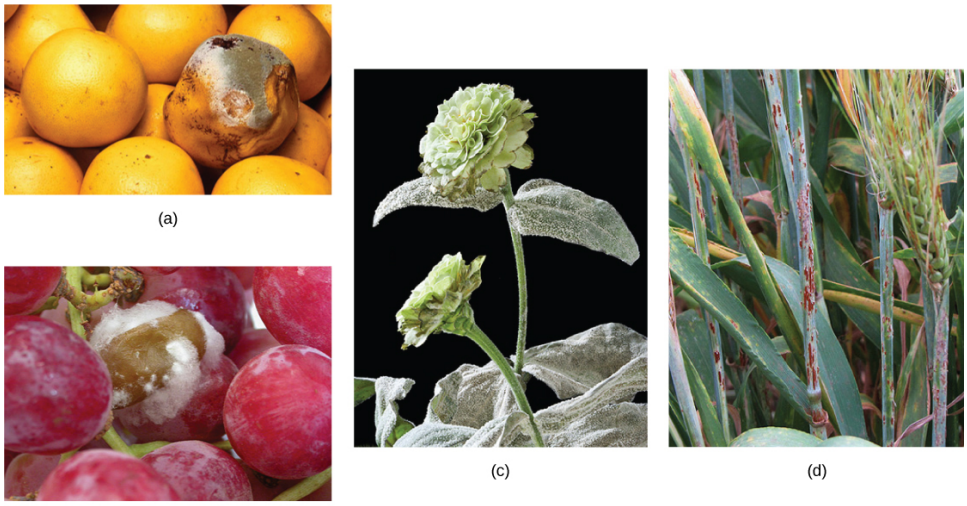


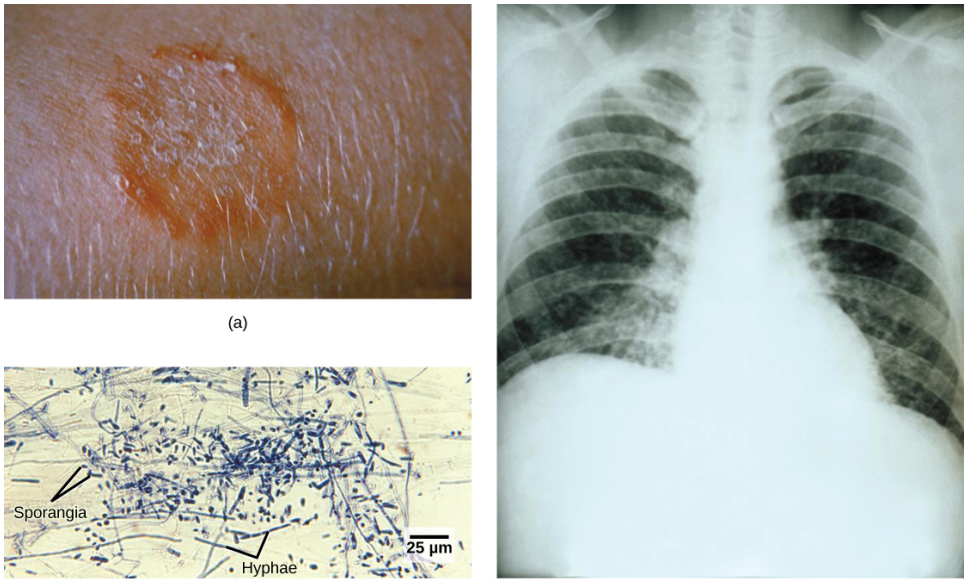
Figure 20.38 Some fungal pathogens include (a) green mold on grapefruit, (b) fungus on grapes, (c) powdery mildew on a zinnia, and (d) stem rust on a sheaf of barley. Notice the brownish color of the fungus in (b) *Botrytis cinerea*, also referred to as the "noble rot," which grows on grapes and other fruit. Controlled infection of grapes by *Botrytis* is used to produce strong and much-prized dessert wines. (credit a: modification of work by Scott Bauer, USDA ARS; credit b: modification of work by Stephen Ausmus, USDA ARS; credit c: modification of work by David Marshall, USDA ARS; credit d: modification of work by Joseph Smilanick, USDA ARS)

Aflatoxins are toxic and carcinogenic compounds released by fungi of the genus *Aspergillus*. Periodically, harvests of nuts and grains are tainted by aflatoxins, leading to massive recall of produce, sometimes ruining producers, and causing food shortages in developing countries.

Animal and Human Parasites and Pathogens

Fungi can affect animals, including humans, in several ways. Fungi attack animals directly by colonizing and destroying tissues. Humans and other animals can be poisoned by eating toxic mushrooms or foods contaminated by fungi. In addition, individuals who display hypersensitivity to molds and spores develop strong and dangerous allergic reactions. Fungal infections are generally very difficult to treat because, unlike bacteria, fungi are eukaryotes. Antibiotics only target prokaryotic cells, whereas compounds that kill fungi also adversely affect the eukaryotic animal host.

Many fungal infections (**mycoses**) are superficial and termed cutaneous (meaning "skin") mycoses. They are usually visible on the skin of the animal. Fungi that cause the superficial mycoses of the epidermis, hair, and nails rarely spread to the underlying tissue (**Figure 20.39**). These fungi are often misnamed "dermatophytes" from the Greek *dermis* skin and *phyte* plant, but they are not plants. Dermatophytes are also called "ringworms" because of the red ring that they cause on skin (although the ring is caused by fungi, not a worm). These fungi secrete extracellular enzymes that break down keratin (a protein found in hair, skin, and nails), causing a number of conditions such as athlete's foot, jock itch, and other cutaneous fungal infections. These conditions are usually treated with over-the-counter topical creams and powders, and are easily cleared. More persistent, superficial mycoses may require prescription oral medications.



(a) (b) (c)

Figure 20.39 (a) Ringworm presents as a red ring on the skin. (b) *Trichophyton violaceum* is a fungus that causes superficial mycoses on the scalp. (c) *Histoplasma capsulatum*, seen in this X-ray as speckling of light areas in the lung, is a species of Ascomycota that infects airways and causes symptoms similar to the flu. (credit a, b: modification of work by Dr. Lucille K. Georg, CDC; credit c: modification of work by M Renz, CDC; scale-bar data from Matt Russell)

Systemic mycoses spread to internal organs, most commonly entering the body through the respiratory system. For example, coccidioidomycosis (valley fever) is commonly found in the southwestern United States, where the fungus resides in the dust. Once inhaled, the spores develop in the lungs and cause signs and symptoms similar to those of tuberculosis. Histoplasmosis (**Figure 20.39c**) is caused by the dimorphic fungus *Histoplasma capsulatum*; it causes pulmonary infections and, in rare cases, swelling of the membranes of the brain and spinal cord. Treatment of many fungal diseases requires the use of antifungal medications that have serious side effects.

Opportunistic mycoses are fungal infections that are either common in all environments or part of the normal biota. They affect mainly individuals who have a compromised immune system. Patients in the late stages of AIDS suffer from opportunistic mycoses, such as *Pneumocystis*, which can be life threatening. The yeast *Candida* spp., which is a common member of the natural biota, can grow unchecked if the pH, the immune defenses, or the normal population of bacteria is altered, causing yeast infections of the vagina or mouth (oral thrush).

Fungi may even take on a predatory lifestyle. In soil environments that are poor in nitrogen, some fungi resort to predation of nematodes (small roundworms). Species of *Arthrobotrys* fungi have a number of mechanisms to trap nematodes. For example, they have constricting rings within their network of hyphae. The rings swell when the nematode touches it and closes around the body of the nematode, thus trapping it. The fungus extends specialized hyphae that can penetrate the body of the worm and slowly digest the hapless prey.

Beneficial Fungi

Fungi play a crucial role in the balance of ecosystems. They colonize most habitats on Earth, preferring dark, moist conditions. They can thrive in seemingly hostile environments, such as the tundra, thanks to a most successful symbiosis with photosynthetic organisms, like lichens. Fungi are not obvious in the way that large animals or tall trees are. Yet, like bacteria, they are major decomposers of nature. With their versatile metabolism, fungi break down organic matter that is insoluble and would not be recycled otherwise.

Importance to Ecosystems

Food webs would be incomplete without organisms that decompose organic matter and fungi are key participants in this process. Decomposition allows for cycling of nutrients such as carbon, nitrogen, and phosphorus back into the environment so they are available to living things, rather than being trapped in dead organisms. Fungi are particularly important because they have evolved enzymes to break down cellulose and lignin, components of plant cell walls that few other organisms are able to digest, releasing their carbon content.

Fungi are also involved in ecologically important coevolved symbioses, both mutually beneficial and pathogenic with organisms from the other kingdoms. **Mycorrhiza**, a term combining the Greek roots *myco* meaning fungus and *rhizo*

meaning root, refers to the association between vascular plant roots and their symbiotic fungi. Somewhere between 80–90 percent of all plant species have mycorrhizal partners. In a mycorrhizal association, the fungal mycelia use their extensive network of hyphae and large surface area in contact with the soil to channel water and minerals from the soil into the plant. In exchange, the plant supplies the products of photosynthesis to fuel the metabolism of the fungus. Ectomycorrhizae ("outside" mycorrhiza) depend on fungi enveloping the roots in a sheath (called a mantle) and a net of hyphae that extends into the roots between cells. In a second type, the Glomeromycota fungi form arbuscular mycorrhiza. In these mycorrhiza, the fungi form arbuscles, a specialized highly branched hypha, which penetrate root cells and are the sites of the metabolic exchanges between the fungus and the host plant. Orchids rely on a third type of mycorrhiza. Orchids form small seeds without much storage to sustain germination and growth. Their seeds will not germinate without a mycorrhizal partner (usually Basidiomycota). After nutrients in the seed are depleted, fungal symbionts support the growth of the orchid by providing necessary carbohydrates and minerals. Some orchids continue to be mycorrhizal throughout their lifecycle.

Lichens blanket many rocks and tree bark, displaying a range of colors and textures. Lichens are important pioneer organisms that colonize rock surfaces in otherwise lifeless environments such as are created by glacial recession. The lichen is able to leach nutrients from the rocks and break them down in the first step to creating soil. Lichens are also present in mature habitats on rock surfaces or the trunks of trees. They are an important food source for caribou. Lichens are not a single organism, but rather a fungus (usually an Ascomycota or Basidiomycota species) living in close contact with a photosynthetic organism (an alga or cyanobacterium). The body of a lichen, referred to as a thallus, is formed of hyphae wrapped around the green partner. The photosynthetic organism provides carbon and energy in the form of carbohydrates and receives protection from the elements by the thallus of the fungal partner. Some cyanobacteria fix nitrogen from the atmosphere, contributing nitrogenous compounds to the association. In return, the fungus supplies minerals and protection from dryness and excessive light by encasing the algae in its mycelium. The fungus also attaches the symbiotic organism to the substrate.

Fungi have evolved mutualistic associations with numerous arthropods. The association between species of Basidiomycota and scale insects is one example. The fungal mycelium covers and protects the insect colonies. The scale insects foster a flow of nutrients from the parasitized plant to the fungus. In a second example, leaf-cutting ants of Central and South America literally farm fungi. They cut disks of leaves from plants and pile them up in gardens. Fungi are cultivated in these gardens, digesting the cellulose that the ants cannot break down. Once smaller sugar molecules are produced and consumed by the fungi, they in turn become a meal for the ants. The insects also patrol their garden, preying on competing fungi. Both ants and fungi benefit from the association. The fungus receives a steady supply of leaves and freedom from competition, while the ants feed on the fungi they cultivate.

Importance to Humans

Although we often think of fungi as organisms that cause diseases and rot food, fungi are important to human life on many levels. As we have seen, they influence the well-being of human populations on a large scale because they help nutrients cycle in ecosystems. They have other ecosystem roles as well. For example, as animal pathogens, fungi help to control the population of damaging pests. These fungi are very specific to the insects they attack and do not infect other animals or plants. The potential to use fungi as microbial insecticides is being investigated, with several species already on the market. For example, the fungus *Beauveria bassiana* is a pesticide that is currently being tested as a possible biological control for the recent spread of emerald ash borer. It has been released in Michigan, Illinois, Indiana, Ohio, West Virginia, and Maryland.

The mycorrhizal relationship between fungi and plant roots is essential for the productivity of farmland. Without the fungal partner in the root systems, 80–90% of trees and grasses would not survive. Mycorrhizal fungal inoculants are available as soil amendments from gardening supply stores and promoted by supporters of organic agriculture.

We also eat some types of fungi. Mushrooms figure prominently in the human diet. Morels, shiitake mushrooms, chanterelles, and truffles are considered delicacies (**Figure 20.40**). The humble meadow mushroom, *Agaricus campestris*, appears in many dishes. Molds of the genus *Penicillium* ripen many cheeses. They originate in the natural environment such as the caves of Roquefort, France, where wheels of sheep milk cheese are stacked to capture the molds responsible for the blue veins and pungent taste of the cheese.

Figure 20.40 The morel mushroom is an ascomycete that is much appreciated for its delicate taste. (credit: Jason Hollinger)

Fermentation—of grains to produce beer, and of fruits to produce wine—is an ancient art that humans in most cultures have practiced for millennia. Wild yeasts are acquired from the environment and used to ferment sugars into CO_2 and ethyl alcohol under anaerobic conditions. It is now possible to purchase isolated strains of wild yeasts from different wine-making regions. Pasteur was instrumental in developing a reliable strain of brewer's yeast, *Saccharomyces cerevisiae*, for the French brewing industry in the late 1850s. It was one of the first examples of biotechnology patenting. Yeast is also used to make breads that rise. The carbon dioxide they produce is responsible for the bubbles produced in the dough that become the air pockets of the baked bread.

Many secondary metabolites of fungi are of great commercial importance. Antibiotics are naturally produced by fungi to kill or inhibit the growth of bacteria, and limit competition in the natural environment. Valuable drugs isolated from fungi include the immunosuppressant drug cyclosporine (which reduces the risk of rejection after organ transplant), the precursors of steroid hormones, and ergot alkaloids used to stop bleeding. In addition, as easily cultured eukaryotic organisms, some fungi are important model research organisms including the red bread mold *Neurospora crassa* and the yeast, *S. cerevisiae.*

21 | EVOLUTION AND DIVERSITY OF PLANTS

21.1 | Lichens, Protists and Green Algae

Introduction

> “ A process which led from the amoeba to man appeared to the philosophers to be obviously a progress—though whether the amoeba would agree with this opinion is not known.”

Bertrand Russell, from "Current Tendencies", delivered as the first of a series of Lowell Lectures in Boston (Mar 1914).

Lichens

Lichens display a range of colors and textures (Figure 21.1) and can survive in the most unusual and hostile habitats. They cover rocks, gravestones, tree bark, and the ground in the tundra where plant roots cannot penetrate. Lichens can survive extended periods of drought, when they become completely desiccated, and then rapidly become active once water is available again.

(a) (b) (c)

Figure 21.1 Lichens have many forms. They may be (a) crust-like, (b) hair-like, or (c) leaf-like. (credit a: modification of work by Jo Naylor; credit b: modification of work by "djpmapleferryman"/Flickr; credit c: modification of work by Cory Zanker)

Lichens are an example of a mutualism, in which a fungus (usually a member of the Ascomycota or Basidiomycota phyla) lives in close contact with a photosynthetic organism (a eukaryotic alga or a prokaryotic cyanobacterium) (Figure 21.2). Generally, neither the fungus nor the photosynthetic organism can survive alone outside of the symbiotic relationship. The body of a lichen, referred to as a thallus, is formed of hyphae wrapped around the photosynthetic partner. The photosynthetic organism provides carbon and energy in the form of carbohydrates. Some cyanobacteria fix nitrogen from the atmosphere, contributing nitrogenous compounds to the association. In return, the fungus supplies minerals and protection from dryness and excessive light by encasing the algae in its mycelium. The fungus also attaches the symbiotic organism to the substrate.

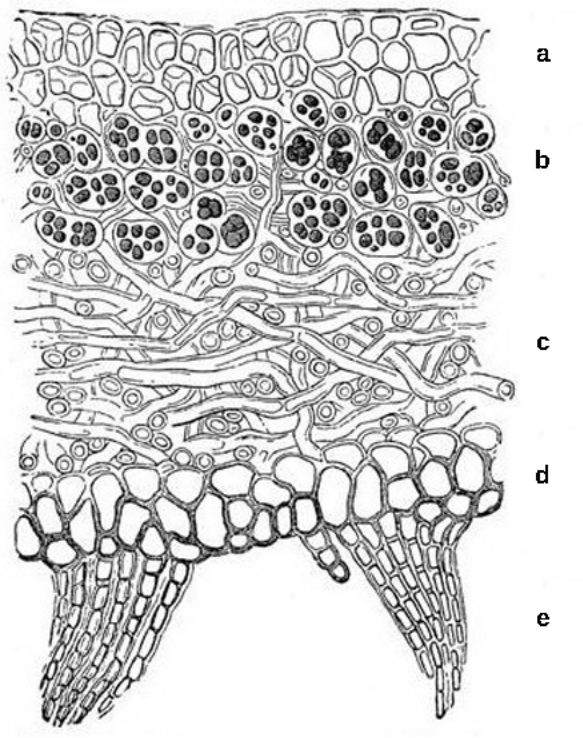


Figure 21.2 This cross-section of a lichen thallus shows the (a) upper cortex of fungal hyphae, which provides protection; the (b) algal zone where photosynthesis occurs, the (c) medulla of fungal hyphae, and the (d) lower cortex, which also provides protection and may have (e) rhizines to anchor the thallus to the substrate.

Lichens grow very slowly, expanding a few millimeters per year. Both the fungus and the alga participate in the formation of dispersal units for reproduction. Lichens produce soredia, clusters of algal cells surrounded by mycelia. Soredia are dispersed by wind and water and form new lichens.

Lichens are extremely sensitive to air pollution, especially to abnormal levels of nitrogen and sulfur. The U.S. Forest Service and National Park Service can monitor air quality by measuring the relative abundance and health of the lichen population in an area. Lichens fulfill many ecological roles. Lichens are often early colonizers of bare rock. Caribou and reindeer eat lichens, and they provide cover for small invertebrates that hide in the mycelium. In the production of textiles, weavers used lichens to dye wool for many centuries until the advent of synthetic dyes.

Protists

Amoebae are just one of the creatures that are lumped into the Kingdom Protista, and amoebae and philosophers do share a common ancestor, as Russell points out. In the span of the last several decades, the Kingdom Protista has been disassembled and rearranged, as DNA sequence analyses have revealed new genetic (and therefore evolutionary) relationships among these eukaryotes. Moreover, protists species that exhibit similar morphological features may not be closely related, but may have evolved analogous structures because of similar selective pressures — rather than because of recent common ancestry. This phenomenon, called **convergent evolution**, is one reason why protist classification is so challenging. The emerging classification scheme groups the entire domain Eukarya into six “supergroups” that contain all of the protists as well as animals, plants, and fungi that evolved from a common ancestor (**Figure 21.3**). The supergroups are hypothesized to be monophyletic, meaning that all organisms within each supergroup are hypothesized to have evolved from a single common ancestor, and thus all members are more closely related to each other than to organisms outside that group. There is still evidence lacking for the monophyly of some groups.

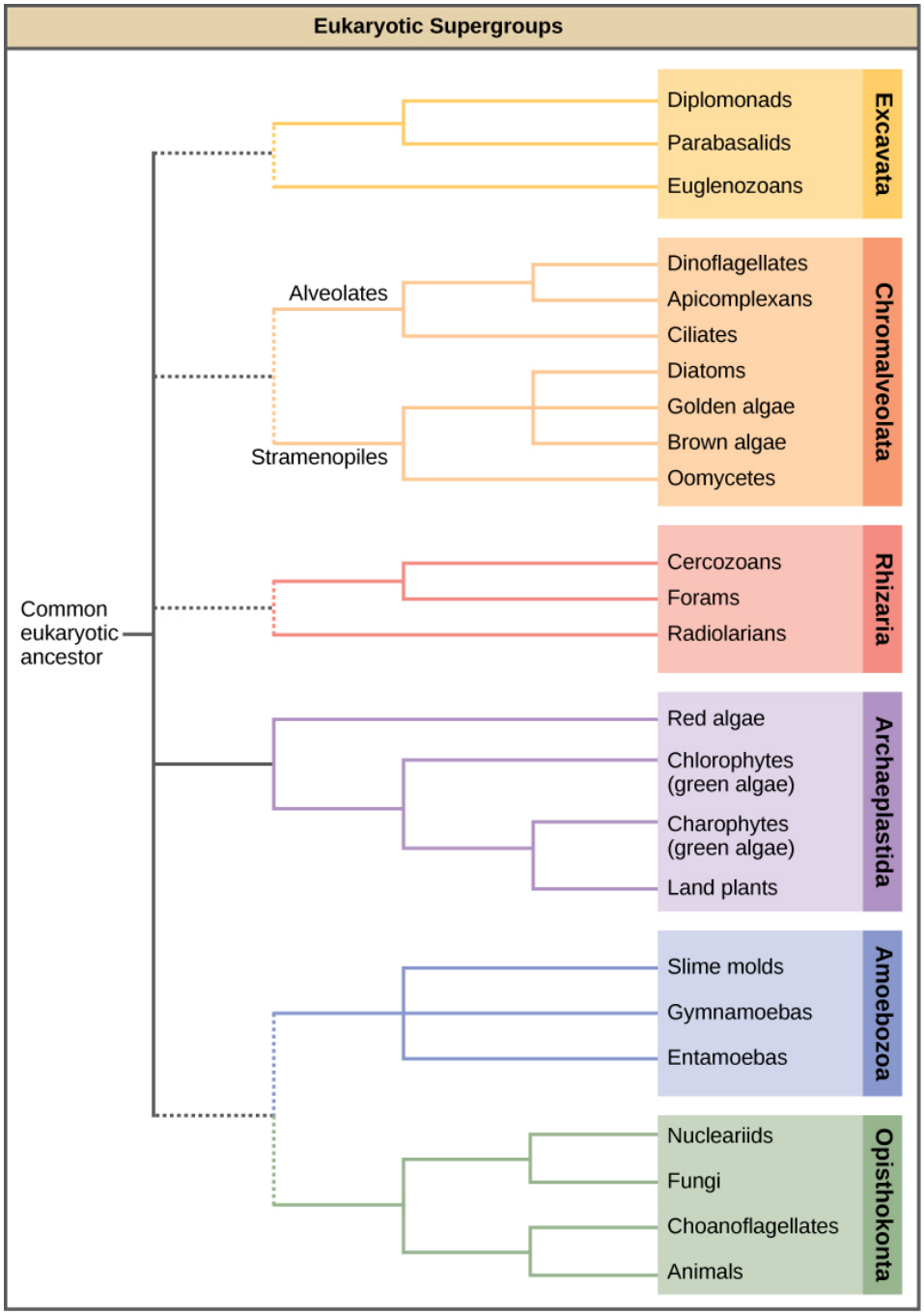


Figure 21.3 This diagram shows a proposed classification of the domain Eukara. Currently, the domain Eukarya is divided into six supergroups. Within each supergroup are multiple kingdoms. Dotted lines indicate suggested evolutionary relationships that remain under debate.

The classification of eukaryotes is still in flux, and the six supergroups may be modified or replaced by a more appropriate hierarchy as genetic, morphological, and ecological data accumulate. Keep in mind that the classification scheme presented here is just one of several hypotheses, and the true evolutionary relationships are still to be determined. For this module, we are focusing on the Archaeplastida which contain the green algae - the group of organisms most closely related to plants.

Archaeplastida (Red Algae, Green Algae and Plants)

Red algae and green algae are included in the supergroup Archaeplastida. It was from a common ancestor of these organisms that the land plants evolved, since their closest relatives are found in this group. Molecular evidence supports that all Archaeplastida are descendents of an endosymbiotic relationship between a heterotrophic protist and a cyanobacterium. The red and green algae include unicellular, multicellular, and colonial forms.

Red Algae

Red algae, or rhodophytes, lack flagella and range in size from microscopic unicellular forms to large, multicellular forms grouped into the informal 'seaweed' category. Most red algae are multicellular. The red algae life cycle is an alternation of generations (explained in next section). Some species of red algae contain phycoerythrins, photosynthetic accessory pigments that are red in color and outcompete the green tint of chlorophyll, making these species appear as varying shades of red. Other species classified as red algae lack phycoerythrins and are parasites. Red algae are common in tropical waters where they have been detected at depths of 260 meters. Other red algae exist in terrestrial or freshwater environments.

Red Algae are an economically important food source and additive. Have you ever eaten sushi rolls? If so, the crispy sheets wrapped around the rice are from the genus Porphyra (Japanese "nori"). The food stabilizer carrageenan is found in ice cream, yogurt and other food stuffs and is an extract from red algae. Another extract from red algae commonly used as a thickener and as vegetarian substitute for gelatin is agar. Agar is also commonly used by microbiologists as a solid substrate to contain culture media in order to grow bacteria.

Green Algae: Chlorophytes and Charophytes

The most abundant group of algae are the green algae. The green algae exhibit similar features to the land plants. The cell walls of green algae and land plants are made of cellulose and the chloroplasts of both groups contain chlorophylls a and b. The hypothesis that this group of protists shared a relatively recent common ancestor with land plants is well supported. The green algae are subdivided into the chlorophytes and charophytes. The charophytes are the closest living relatives to land plants and resemble them in morphology and reproductive strategies. Charophytes are common in wet habitats, and their presence often signals a healthy ecosystem.

Green algae are as a group of organisms an integral part of a functional ecosystem, and humans have been using green algae as a food source and as a medicine for a long time. In aquatic environments, green algae are a major primary producer and release a substantial amount of oxygen in the system. So a healthy ecosystem is dependent on a healthy population of green algae. As a economic benefit to humans, green algae are used a food source (Asakusa Nori) for humans and are used a food thickening agent. Green algae are used in the agricultural industry as a food source for cattle, and as fertilizer.

The chlorophytes exhibit great diversity of form and function. Chlorophytes primarily inhabit freshwater and damp soil, and are a common component of plankton. *Chlamydomonas* is a simple, unicellular chlorophyte with a pear-shaped morphology and two opposing, anterior flagella that guide this protist toward light sensed by its eyespot.

Volvox is an example of multicellularity in the Chlorophytes (**Figure 21.4**). *Volvox* colonies contain 500 to 60,000 cells, each with two flagella, contained within a hollow, spherical matrix composed of a gelatinous glycoprotein secretion. *Volvox* moves by rolling in a coordinated fashion. The cells forming the sphere on the outside do not reproduce while the green cells inside do reproduce, demonstrating division of labor.

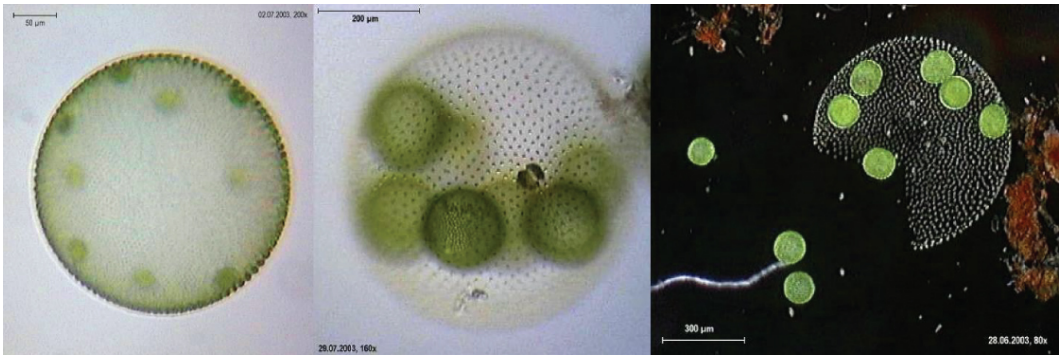


Figure 21.4 *Volvox aureus* is a green alga in the supergroup Archaeplastida. This species exists as a colony, consisting of cells immersed in a gel-like matrix and intertwined with each other via hair-like cytoplasmic extensions. (credit: Dr. Ralf Wagner)

True multicellular organisms, such as the sea lettuce, *Ulva*, are represented among the chlorophytes. In addition, some chlorophytes exist as large, multinucleate, single cells. Species in the genus *Caulerpa* exhibit flattened fern-like foliage and can reach lengths of 3 meters (**Figure 21.5**). *Caulerpa* species undergo nuclear division, but their cells do not complete cytokinesis, remaining instead as massive and elaborate single cells.

Figure 21.5 *Caulerpa taxifolia* is a chlorophyte consisting of a single cell containing potentially thousands of nuclei. (credit: NOAA)

Charophytes are the closest related green algae to land plants, this conclusion was drawn from studies of nuclear and chloroplast genes from many different types of plants and algae. In addition to the similarities listed above, there a some additional structural similarities such as cellulose-synthesizing proteins, the structure of the flagellated sperm and the gamete producing structures that suggest that land plants arose from within the charophyta group. This does not mean land plants are descended from living green algae, but it does allow us to explore the algal ancestor to land plants. For example, the exploration of the Chara life cycle **Figure 21.6** allows us visualize the similarities between the life cycle of green algae and land plants.

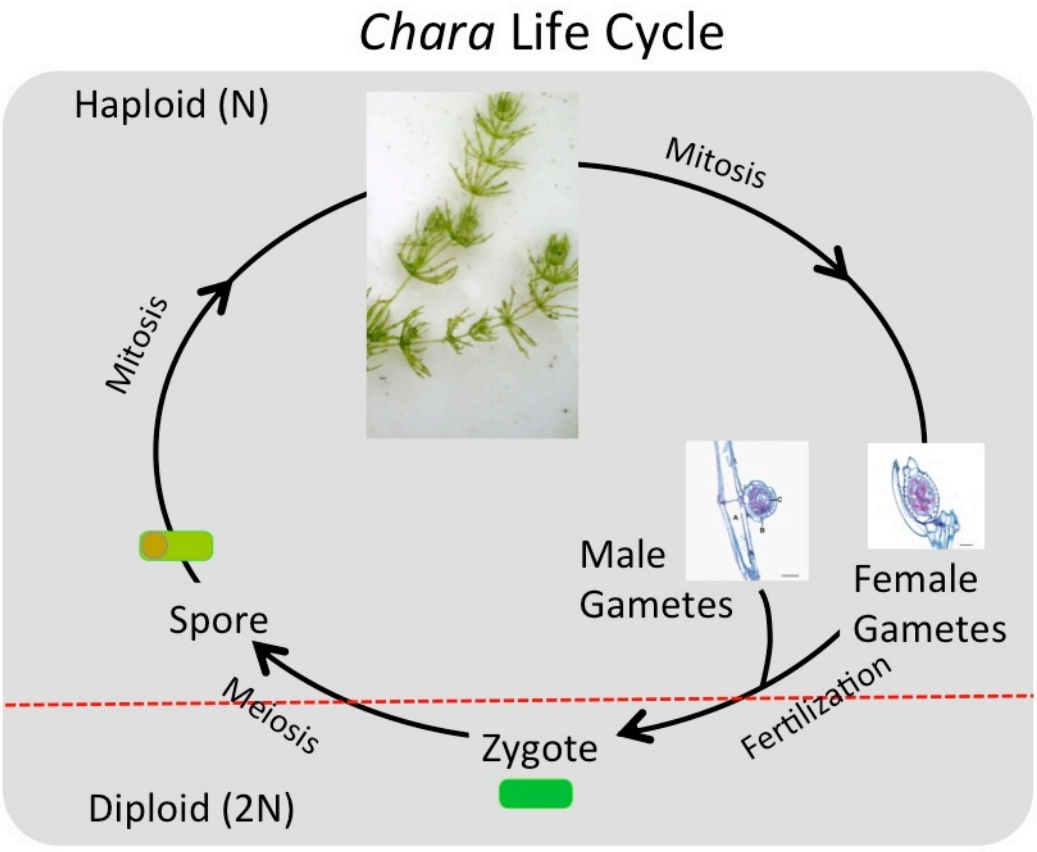


Figure 21.6 Simplified life cycle of Chara. (Work by Robbie Bear, Images credited to "Chara antheridia L" by Jon Houseman - Jon Houseman and Matthew Ford. Licensed under CC BY-SA 4.0 via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Chara_antheridia_L.jpg#/media/File:Chara_antheridia_L.jpg "Chara oogonium" by Jon Houseman - Jon Houseman and Matthew Ford. Licensed under CC BY-SA 4.0 via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Chara_oogonium.jpg#/media/File:Chara_oogonium.jpg "Chara braunii 1" by Show_ryu - Own work. Licensed under CC BY-SA 3.0 via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Chara_braunii_1.JPG#/media/File:Chara_braunii_1.JPG)

21.2 | Early Plant Life

Introduction

> “I observed on most collected stones the imprints of innumerable plant fragments which were so different from those which are growing in the Lyonnais, in the nearby provinces, and even in the rest of France, that I felt like collecting plants in a new world... The number of these leaves, the way they separated easily, and the great variety of plants whose imprints I saw, appeared to me just as many volumes of botany representing in the same quarry the oldest library of the world. ”

Antoine de Jussieu, French physician and botanist, 1718

The kingdom Plantae constitutes a large and varied group of organisms, which have been on the planet for a very long time. There are more than 300,000 species of catalogued plants, including the fossil plants that de Jussieu references in the epigraph above. Of these, more than 260,000 are seed plants. Mosses, ferns, conifers, and flowering plants are all members of the plant kingdom. While there is some disagreement about the relationships between Chlorophytes, Charophytes, and Plantae, there are several unique characteristics which these groups share. Only green algae and plants use chlorophyll a and b plus carotene in a particular ratio. They share the trait of cellulose-rich cell walls, and there is strong molecular support for their close relationship.

The ancestors of the green algae became photosynthetic by engulfing a green, photosynthetic bacterium about 1.65 billion years ago. This captured bacterium evolved into a chloroplast. That algal line evolved into the Charophytes, bryophytes, seedless vascular plants, gymnosperms, and angiosperms.

Several other groups of eukaryotes have common names that include ‘algae.’ In the latest classification red algae are included in Archaeplastida, while brown algae and golden algae are in a separate supergroup. In contrast to the green algae, red, golden, and brown algae all became photosynthetic by secondary, or even tertiary, endosymbiotic events. In other words, the cells that evolved into red, brown, or golden algae all engulfed cells that had already engulfed a photosynthetic bacterium. These algae are also photosynthetic autotrophs, but they did not diversify to the same extent as the Charophytes, nor did they colonize land.

Plant Adaptations to Life on Land

In order for plants to invade land, they had to contend with several challenges in the terrestrial environment. Water has been described as “the stuff of life.” The cell’s interior is a watery soup: in this medium, most small molecules can dissolve and diffuse rapidly, and the majority of the chemical reactions of metabolism take place. The first challenge, **Desiccation**, or drying out, is a constant danger for an organism exposed to air. Even when parts of a plant are close to a source of water, the aerial structures are likely to dry out. Second, Water also provides buoyancy to organisms. On land, plants need to develop structural support in a medium that does not give the same support as water. The organism is also subject to bombardment by mutagenic radiation, because air does not filter out ultraviolet rays of sunlight like water does. Additionally, the male gametes must reach the female gametes using new strategies, because swimming is no longer possible. Lastly, both gametes and zygotes must be protected from desiccation. The successful land plants developed strategies to deal with all of these challenges. Not all adaptations appeared at once. Some species never moved very far from the aquatic environment, whereas others went on to conquer the driest environments on Earth.

To balance these survival challenges, life on land offers several advantages. First, sunlight is abundant. Water acts as a filter, altering the spectral quality of light absorbed by the photosynthetic pigment chlorophyll. Second, carbon dioxide is more readily available in air than in water, since it diffuses faster in air. Third, land plants evolved before land animals; therefore, until dry land was colonized by animals, no predators threatened plant life. This situation changed as animals emerged from the water and fed on the abundant sources of nutrients in the established flora. In turn, plants developed strategies to deter predation: from spines and thorns to toxic chemicals.

Early land plants, like the early land animals, did not live very far from an abundant source of water and developed survival strategies to combat dryness. One of these strategies is called tolerance. Many mosses, for example, can dry out to a brown and brittle mat, but as soon as rain or a flood makes water available, mosses will absorb it and are restored to their healthy green appearance. Another strategy is to colonize environments with high humidity, where droughts are uncommon. Ferns, which are considered an early lineage of plants, thrive in damp and cool places such as the understory of temperate forests. Later, plants moved away from moist or aquatic environments using resistance to desiccation, rather than tolerance. These plants, like cacti, minimize the loss of water to such an extent they can survive in extremely dry environments.

The most successful adaptation was the development of new structures that gave plants the advantage when colonizing new and dry environments. Four major adaptations are found in all terrestrial plants: the alternation of generations, a sporangium in which the spores are formed, a gametangium that produces haploid cells, and apical meristem tissue in roots and shoots. The evolution of a waxy cuticle and a cell wall with lignin also contributed to the success of land plants. These adaptations are noticeably lacking in the closely related green algae.

Alternation of Generations

Alternation of generations describes a life cycle in which an organism has both haploid and diploid multicellular stages (**Figure 21.7**).

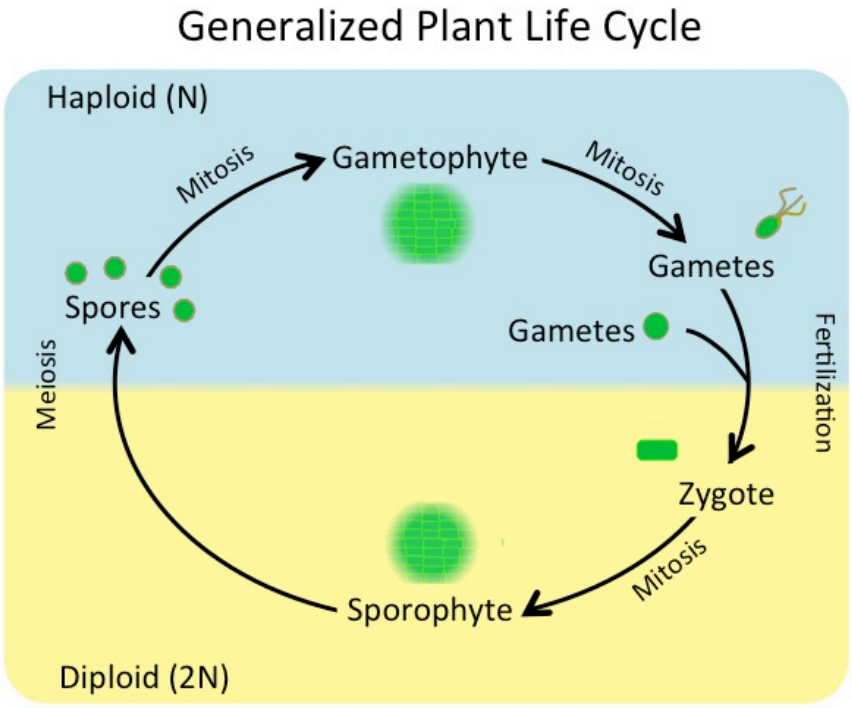


Figure 21.7 Alternation of generations between the 1N gametophyte and 2N sporophyte is shown. Work by Eva Horne and Robert A. Bear

Most plants exhibit alternation of generations, wherein the haploid multicellular form, known as a **gametophyte**, is followed in the development sequence by a multicellular diploid organism: the **sporophyte**. The gametophyte gives rise to the gametes (reproductive cells) by mitosis. This can be the most obvious phase of the life cycle of the plant, as in the mosses, or it can occur in a microscopic structure, such as a pollen grain, in the higher plants (a common collective term for the vascular plants). The sporophyte stage is barely noticeable in lower plants (the collective term for the plant groups of mosses, liverworts, and hornworts). Towering trees are the diploid, sporophyte phase in the lifecycles of plants such as sequoias and pines.

Protection of the embryo is a major requirement for land plants. Embryos have a high surface area to volume ratio, and thus are vulnerable to desiccation and other environmental hazards. In both seedless and seed plants, the female gametophyte provides protection and nutrients to the embryo as it develops into the new generation of sporophyte. This distinguishing feature of land plants gave the group its alternate name of embryophytes.

Sporangia in Seedless Plants

The sporophyte of seedless plants is diploid and results from syngamy (fusion) of two gametes. The sporophyte bears the sporangia (singular, sporangium): organs that first appeared in the land plants. The term “sporangia” literally means “spore in a vessel,” as it is a reproductive sac that contains spores **Figure 21.8**. Inside the multicellular sporangia, the diploid sporocytes, or mother cells, produce haploid spores by meiosis, where the 2N chromosome number is reduced to 1N (note that many plant sporophytes are polyploid: for example, durum wheat is tetraploid (4N), bread wheat is hexaploid (6N), and some ferns are 1000-ploid). The spores are later released by the sporangia and disperse in the environment. Two different types of spores are produced in land plants, resulting in the separation of sexes at different points in the lifecycle. **Seedless non-vascular plants** produce only one kind of spore and are called **homosporous**. The gametophyte phase is dominant in these plants. After germinating from a spore, the resulting gametophyte produces both male and female gametangia, usually on the same individual. In contrast, **heterosporous** plants produce two morphologically different types of spores. The male spores are called **microspores**, because of their smaller size, and develop into the male gametophyte; the comparatively larger **megaspores** develop into the female gametophyte. Heterospory is observed in a few **seedless vascular plants** and in all seed plants.

Figure 21.8 Spore-producing sacs called sporangia grow at the ends of long, thin stalks in this photo of the moss *Esporangios bryum*. (credit: Javier Martin)

When the haploid spore germinates in a hospitable environment, it grows into a multicellular gametophyte by mitosis. The gametophyte supports the zygote formed from the fusion of gametes and the resulting young sporophyte (vegetative form). The cycle then begins anew.

The spores of seedless plants are surrounded by thick cell walls containing a tough polymer known as sporopollenin. This complex substance is characterized by long chains of organic molecules related to fatty acids and carotenoids: hence the yellow color of most pollen. Sporopollenin is unusually resistant to chemical and biological degradation. In seed plants, which use pollen (the microspore) to transfer the male sperm to the female egg, the toughness of sporopollenin explains the existence of well-preserved pollen fossils. Sporopollenin was once thought to be an innovation of land plants; however, the green algae *Coleochaetes* forms spores that contain sporopollenin.

Gametangia in Seedless Plants

Gametangia (singular, gametangium) are structures observed on multicellular haploid gametophytes. In the gametangia, precursor cells give rise to gametes by mitosis. The male gametangium (antheridium) releases sperm. Many seedless plants produce sperm equipped with flagella that enable them to swim in a moist environment to the archegonia: the female gametangium. The embryo develops inside an archegonium into a multicellular sporophyte. Gametangia are prominent in seedless plants, but are very rarely found in seed plants.

Apical Meristems

Shoots and roots of plants increase in length through rapid cell division in a tissue called the apical meristem, which is a small zone of cells found at the shoot tip or root tip (**Figure 21.9**). The apical meristem is made of undifferentiated cells that continue to proliferate throughout the life of the plant. Meristematic cells give rise to all the specialized tissues of the organism. Elongation of the shoots and roots allows a plant to access additional space and resources: light in the case of the shoot, and water and minerals in the case of roots. A separate meristem, called the lateral meristem, produces cells that increase the diameter of tree trunks.

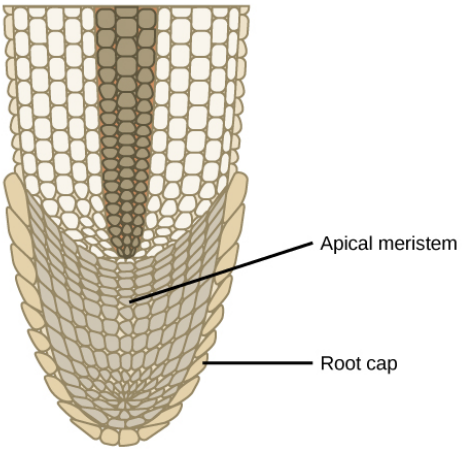


Figure 21.9 Addition of new cells in a root occurs at the apical meristem. Subsequent enlargement of these cells causes the organ to grow and elongate. The root cap protects the fragile apical meristem as the root tip is pushed through the soil by cell elongation.

Additional Land Plant Adaptations

As plants adapted to dry land and became independent from the constant presence of water in damp habitats, new organs and structures made their appearance. Early land plants did not grow more than a few inches off the ground, competing for light on these low mats. By developing a shoot and growing taller, individual plants captured more light. Because air offers substantially less support than water, land plants incorporated more rigid molecules in their stems (and later, tree trunks). In small organisms such as single-celled algae, simple diffusion suffices to distribute water and nutrients throughout the organism. However, for plants to evolve larger forms, the evolution of vascular tissue for the distribution of water and solutes was a prerequisite. The vascular system contains xylem and phloem tissues. **Xylem** conducts water and minerals absorbed from the soil up to the shoot, while **phloem** transports food derived from photosynthesis throughout the entire plant. A root system evolved to take up water and minerals from the soil, and to anchor the increasingly taller shoot in the soil.

In land plants, a waxy, waterproof cover called a cuticle protects the leaves and stems from desiccation. However, the cuticle also prevents intake of carbon dioxide needed for the synthesis of carbohydrates through photosynthesis. To overcome this, stomata or pores that open and close to regulate traffic of gases and water vapor evolved in plants as they moved away from moist environments into drier habitats.

Lignin is a complex polymer predominantly found in the cell walls of xylem, it forms crosslinks between the cellulose molecules. Since lignin is interwoven into the cell walls, it adds strength to the cell wall and therefore the entire plant. In addition, lignin is a hydrophobic compound that allows for the more efficient transport of water in the vascular tissue.

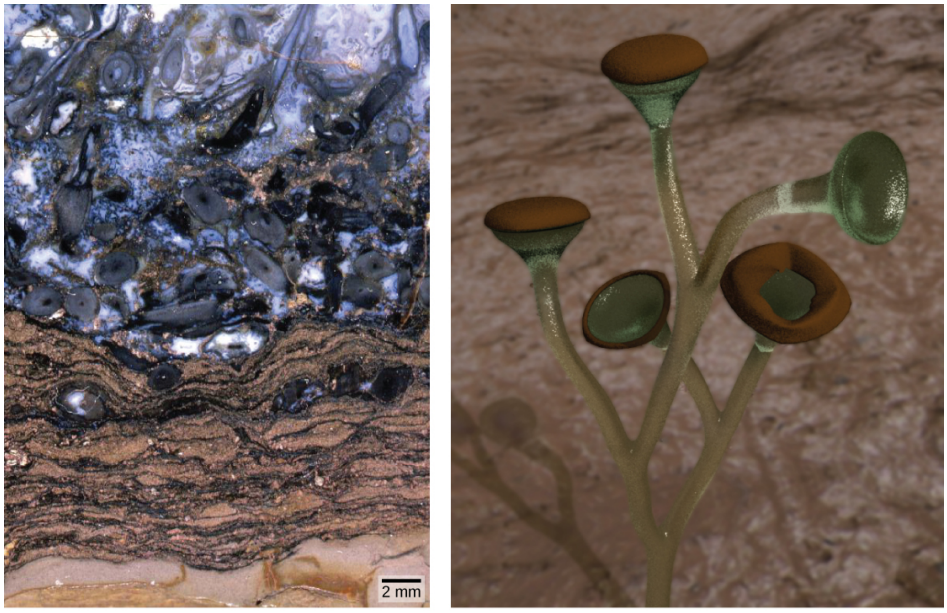
Water filters ultraviolet-B (UVB) light, which is harmful to all organisms, especially those that must absorb light to survive. This filtering does not occur for land plants. This presented an additional challenge to land colonization, which was met by the evolution of biosynthetic pathways for the synthesis of protective flavonoids and other compounds: pigments that absorb UV wavelengths of light and protect the aerial parts of plants from photodynamic damage.

Plants cannot avoid being eaten by animals. Instead, they synthesize a large range of poisonous secondary metabolites: complex organic molecules such as alkaloids, whose noxious smells and unpleasant taste deter animals. These toxic compounds can also cause severe diseases and even death, thus discouraging predation. Humans have used many of these compounds for centuries as drugs, medications, or spices. In contrast, as plants co-evolved with animals, the development of sweet and nutritious metabolites lured animals into providing valuable assistance in dispersing pollen grains, fruit, or seeds. Plants have been enlisting animals to be their helpers in this way for hundreds of millions of years.

Evolution of Land Plants

No discussion of the evolution of plants on land can be undertaken without a brief review of the timeline of the geological eras. The early era, known as the Paleozoic, is divided into six periods. It starts with the Cambrian period, followed by the Ordovician, Silurian, Devonian, Carboniferous, and Permian. The major event to mark the Ordovician, more than 500 million years ago, was the colonization of land by the ancestors of modern land plants. Fossilized cells, cuticles, and spores of early land plants have been dated as far back as the Ordovician period in the early Paleozoic era. The oldest-known vascular plants have been identified in deposits from the Devonian. One of the richest sources of information is the Rhynie chert, a sedimentary rock deposit found in Rhynie, Scotland (**Figure 21.10**), where embedded fossils of some of the earliest

vascular plants have been identified.



(a) (b)

Figure 21.10 This Rhynie chert contains fossilized material from vascular plants. The area inside figure a contains bulbous underground stems called corms, and root-like structures called rhizoids. (credit b: modification of work by Peter Coxhead based on original image by “Smith609”/Wikimedia Commons; scale-bar data from Matt Russell)

Paleobotanists distinguish between **extinct** species, as fossils, and **extant** species, which are still living. The extinct vascular plants, classified as zosterophylls and trimerophytes, most probably lacked true leaves and roots and formed low vegetation mats similar in size to modern-day mosses, although some trimetophytes could reach one meter in height. The later genus *Cooksonia*, which flourished during the Silurian, has been extensively studied from well-preserved examples. Imprints of *Cooksonia* show slender branching stems ending in what appear to be sporangia. From the recovered specimens, it is not possible to establish for certain whether *Cooksonia* possessed vascular tissues. Fossils indicate that by the end of the Devonian period, ferns, horsetails, and seed plants populated the landscape, giving rising to trees and forests. This luxuriant vegetation helped enrich the atmosphere in oxygen, making it easier for air-breathing animals to colonize dry land. Plants also established early symbiotic relationships with fungi, creating mycorrhizae: a relationship in which the fungal network of filaments increases the efficiency of the plant root system, and the plants provide the fungi with byproducts of photosynthesis.

career CONNECTION

Paleobotanist

How organisms acquired traits that allow them to colonize new environments—and how the contemporary ecosystem is shaped—are fundamental questions of evolution. Paleobotany (the study of extinct plants) addresses these questions through the analysis of fossilized specimens retrieved from field studies, reconstituting the morphology of organisms that disappeared long ago. Paleobotanists trace the evolution of plants by following the modifications in plant morphology: shedding light on the connection between existing plants by identifying common ancestors that display the same traits. This field seeks to find transitional species that bridge gaps in the path to the development of modern organisms. Fossils are formed when organisms are trapped in sediments or environments where their shapes are preserved. Paleobotanists collect fossil specimens in the field and place them in the context of the geological sediments and other fossilized organisms surrounding them. The activity requires great care to preserve the integrity of the delicate fossils and the layers of rock in which they are found.

One of the most exciting recent developments in paleobotany is the use of analytical chemistry and molecular biology to study fossils. Preservation of molecular structures requires an environment free of oxygen, since oxidation and degradation of material through the activity of microorganisms depend on its presence. One example of the use of analytical chemistry and molecular biology is the identification of oleanane, a compound that deters pests. Up to this point, oleanane appeared to be unique to flowering plants; however, it has now been recovered from sediments dating from the Permian, much earlier than the current dates given for the appearance of the first flowering plants. Paleobotanists can also study fossil DNA, which can yield a large amount of information, by analyzing and comparing the DNA sequences of extinct plants with those of living and related organisms. Through this analysis, evolutionary relationships can be built for plant lineages.

Paleobotanists must be cautious when drawing conclusions from the analysis of molecular fossils. Chemical materials of interest degrade rapidly when exposed to air during their initial isolation, as well as in further manipulations. There is always a risk of contaminating the specimens with extraneous material, mostly from microorganisms. Nevertheless, as technology is refined, the analysis of DNA from fossilized plants will provide invaluable information on the evolution of plants and their adaptation to an ever-changing environment.

The Major Divisions of Land Plants

Land plants are classified into two major groups according to the absence or presence of vascular tissue, as detailed in **Figure 21.11**. Plants that lack vascular tissue, which is formed of specialized cells for the transport of water and nutrients, are referred to as **non-vascular plants**. Liverworts, mosses, and hornworts are seedless, non-vascular plants that likely appeared early in land plant evolution. Vascular plants developed a network of cells that conduct water and solutes. The first vascular plants appeared in the middle Paleozoic and were probably similar to lycophytes, which include club mosses (not to be confused with the mosses) and the pterophytes (ferns, horsetails, and whisk ferns). Lycophytes and pterophytes are referred to as seedless vascular plants, because they do not produce seeds. The seed plants, or spermatophytes, form the largest group of all existing plants, and hence dominate the landscape. Seed plants include Gymnosperms, most notably conifers (Gymnosperms), which produce “naked seeds,” and the most successful of all plants, the flowering plants (Angiosperms). Angiosperms protect their seeds inside chambers at the center of a flower; the walls of the chamber later develop into a fruit.

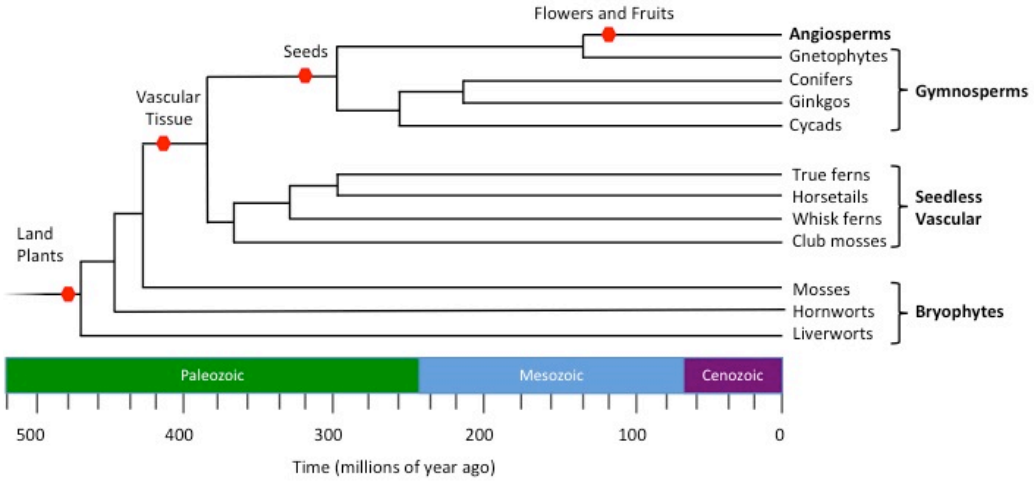


Figure 21.11 This figure shows the major divisions of green plants and when various adaptations for land life evolved. Work by Robert A. Bear

21.3 | Bryophytes

Introduction

> "It was to Hofmeister, working as a young man, an amateur and enthusiast, in the early morning hours of summer months, before business, at Leipzig in the years before 1851, that the vision first appeared of a common type of Life-Cycle, running through Mosses and Ferns to Gymnosperms and Flowering Plants, linking the whole series in one scheme of reproduction and life-history."

Arthur Harry Church, 1919. As quoted in E.J.H. Corner, *The Life of Plants* (1964)

The commonality of the plant life cycle unites the Plant Kingdom, indicating that it appeared very early in the evolution of plants. The Bryophytes are the group of plants that are the closest extant relative of those early terrestrial plants. The first bryophytes (liverworts) most likely appeared in the Ordovician period, about 450 million years ago. Because of the lack of lignin and other resistant structures, the likelihood of bryophytes forming fossils is rather small. Some spores protected by sporopollenin have survived and are attributed to early bryophytes. By the Silurian period, however, vascular plants had spread through the continents. This compelling fact is used as evidence that non-vascular plants must have preceded the Silurian period.

More than 25,000 species of bryophytes thrive in mostly damp habitats, although some live in deserts. They constitute the major flora of inhospitable environments like the tundra, where their small size and tolerance to desiccation offer distinct advantages. They generally lack lignin and do not have actual tracheids (xylem cells specialized for water conduction). Rather, water and nutrients circulate inside specialized conducting cells. Although the term non-tracheophyte is more accurate, bryophytes are commonly called non-vascular plants.

In a bryophyte, all the conspicuous vegetative organs—including the photosynthetic leaf-like structures, the thallus, stem, and the rhizoid that anchors the plant to its substrate—belong to the haploid organism or gametophyte. The sporophyte is barely noticeable. The gametes formed by bryophytes swim with a flagellum, as do gametes in a few of the vascular plants.

The sporangium—the multicellular sexual reproductive structure—is present in bryophytes and absent in the majority of algae. The bryophyte embryo also remains attached to the parent plant, which protects and nourishes it. This is a characteristic of land plants.

The Bryophyte life cycle follows the pattern of alternation of generations as shown in **m47388 (http://legacy.cnx.org/content/m47388/1.6/#bryophyte_lifecycle)** . The most familiar structure is the haploid gametophyte, which germinates from a haploid spore. Cells similar to an apical meristem actively divide and give rise to the photosynthetic stem and leaf-like structures. Sperm and egg producing structures form on separate or the same stems. Sperm swim along the bryophte and unite with the egg inside the egg-producing structure (archegonium). The zygote, protected by this structure, divides and grows into a sporophyte, still attached to the gametophyte. The sporophyte forms spores by meiosis; these disperse and will form new gametophytes.

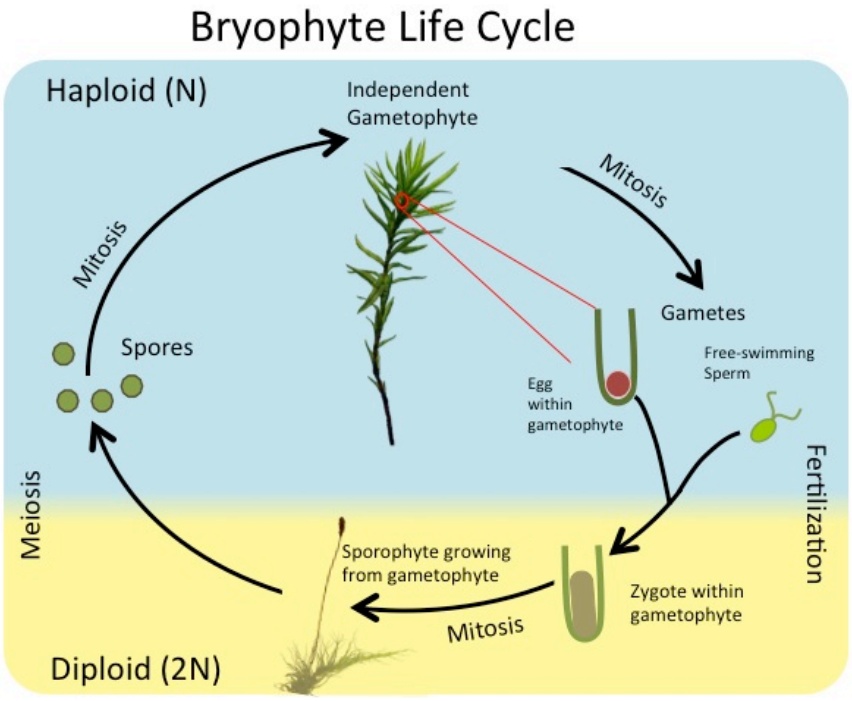


Figure 21.12 This illustration shows the generalized life cycle of bryophytes. Work by Eva Horne and Robert A. Bear

The bryophytes are divided into three phyla: the liverworts or Hepaticophyta, the hornworts or Anthocerotophyta, and the mosses or true Bryophyta. The organisms in these three phyla share the following characteristics. Bryophytes lack vascular tissue, lack true leaves, lack seeds, use spores as a means of dispersal and have the gametophyte generation as the dominant (conspicuous) part of the life cycle. Even with all these characteristics in common, molecular and other evidence suggests that they do not form a single clade (a group that includes one common ancestor and all of its descendants)

Liverworts

Liverworts (Hepaticophyta) are viewed as the plants most closely related to the ancestor that moved to land. Liverworts have colonized every terrestrial habitat on Earth and diversified to more than 7,000 existing species (**Figure 21.13**). Some gametophytes form lobate green structures, as seen in **Figure 21.14**. The shape is similar to the lobes of the liver, and hence provides the origin of the name given to the phylum. Openings that allow the movement of gases may be observed in liverworts. However, these openings are not stomata, because they do not actively open and close. The plant takes up water over its entire surface and has no cuticle to prevent desiccation.

Figure 21.13 This 1904 drawing shows the variety of forms of liverworts.

Figure 21.14 A liverwort, *Lunularia cruciata*, displays its lobate, flat thallus. The organism in the photograph is in the gametophyte stage.

The lifecycle of a liverwort starts with the release of haploid spores from the sporangium that developed on the sporophyte. Spores disseminated by wind or water germinate into flattened thalli attached to the substrate by thin, single-celled filaments. Male and female gametangia develop on separate, individual plants. Once released, male gametes swim with the aid of their flagella to the female gametangium (the archegonium), and fertilization ensues. The zygote grows into a small sporophyte still attached to the parent gametophyte. It will give rise, by meiosis, to the next generation of spores. Liverwort plants can also reproduce asexually, by the breaking of branches or the spreading of leaf fragments called gemmae. In this latter type of reproduction, the gemmae—small, intact, complete pieces of plant that are produced in a cup on the surface of the thallus are splashed out of the cup by raindrops. The gemmae then land nearby and develop into gametophytes.

Hornworts

The **hornworts** (*Anthocerotophyta*) have colonized a variety of habitats on land, although they are never far from a source

of moisture. The short, blue-green gametophyte is the dominant phase of the lifecycle of a hornwort. The narrow, pipe-like sporophyte is the defining characteristic of the group. The sporophytes emerge from the parent gametophyte and continue to grow throughout the life of the plant (**Figure 21.15**).

Figure 21.15 Hornworts grow a tall and slender sporophyte. (credit: modification of work by Jason Hollinger)

Stomata appear in the hornworts and are abundant on the sporophyte. Photosynthetic cells in the thallus contain a single chloroplast. Meristem cells at the base of the plant keep dividing and adding to its height. Many hornworts establish symbiotic relationships with cyanobacteria that fix nitrogen from the environment.

The lifecycle of hornworts follows the general pattern of alternation of generations. The gametophytes grow as flat thalli on the soil with embedded gametangia. Flagellated sperm swim to the archegonia and fertilize eggs. The zygote develops into a long and slender sporophyte that eventually splits open, releasing spores. The haploid spores germinate and give rise to the next generation of gametophyte.

Mosses

More than 10,000 species of **mosses** have been catalogued. Their habitats vary from the tundra, where they are the main vegetation, to the understory of tropical forests. In the tundra, the mosses' shallow rhizoids allow them to fasten to a substrate without penetrating the frozen soil. Mosses slow down erosion, store moisture and soil nutrients, and provide shelter for small animals as well as food for larger herbivores, such as the musk ox. Mosses are very sensitive to air pollution and are used to monitor air quality. They are also sensitive to copper salts, so these salts are a common ingredient of compounds marketed to eliminate mosses from lawns.

Mosses form diminutive gametophytes, which are the dominant phase of the lifecycle. Green, flat structures—resembling true leaves, but lacking vascular tissue—are attached in a spiral to a central stalk. The plants absorb water and nutrients directly through these leaf-like structures. Some mosses have small branches. Some primitive traits of green algae, such as flagellated sperm, are still present in mosses that are dependent on water for reproduction. Other features of mosses are clearly adaptations to dry land. For example, stomata are present on the stems of the sporophyte, and a primitive vascular system runs up the sporophyte's stalk. Additionally, mosses are anchored to the substrate—whether it is soil, rock, or roof tiles—by multicellular rhizoids. These structures are precursors of roots. They originate from the base of the gametophyte, but are not the major route for the absorption of water and minerals. The lack of a true root system explains why it is so easy to rip moss mats from a tree trunk. The moss lifecycle follows the pattern of alternation of generations as shown in **m47388 (http://legacy.cnx.org/content/m47388/1.6/#bryophyte_lifecycle)** .

21.4 | Seedless Vascular Plants

Introduction

> "We might expect... in the summer of the 'great year,' which we are now considering, that there would be a great predominance of tree-ferns and plants allied to the palms and arborescent grasses

> in the isles of the wide ocean, while the dicotyledenous plants and other forms now most common in temperate regions would almost disappear from the earth. Then might these genera of animals return, of which the memorials are preserved in the ancient rocks of our continents. The huge iguanodon might reappear in the woods, and the ichthyosaur in the sea, while the pterodactyle might flit again through umbrageous groves of tree-ferns. Coral reefs might be prolonged beyond the arctic circle, where the whale and narwal [sic] now abound. Turtles might deposit their eggs in the sand of the sea beach, where now the walrus sleeps, and where the seal is drifted on the ice-floe. ”

Sir Charles Lyell, *Principles of Geology*, (1830)

Lyell's book, which greatly influenced scientists in the time of Darwin, appeared just as science was comprehending the value of the fossil record in reconstructing the history of life on earth. Fossilized tree-ferns from coal mines in Britain and North America gave hints of an ancient time of lush tropical forests in those regions, and how and when those forests disappeared was a mystery waiting to be solved. The vascular plants, including ferns and their allies, are the dominant and most conspicuous group of land plants. More than 260,000 species (including ferns, gymnosperms and angiosperms) represent more than 90 percent of Earth’s vegetation. Several evolutionary innovations explain their success and their ability to spread to all habitats.

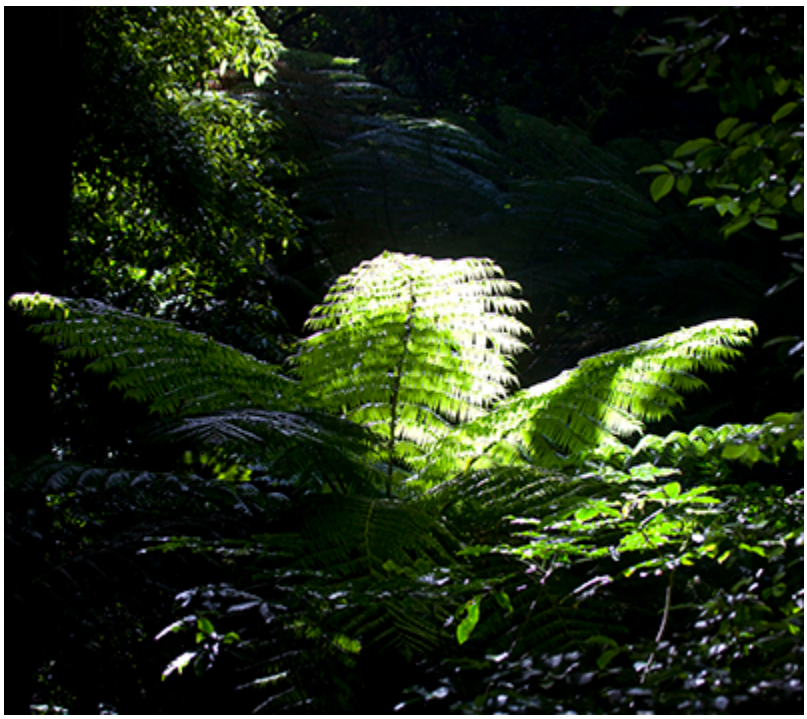


Figure 21.16 A tree-fern in New Zealand A modern tree-fern in the forests of New Zealand. This particular species is the Silver Fern (*Cyathea dealbata*), the national symbol of New Zealand, which appears on the uniforms of the national Rugby Team, the All Blacks.

Bryophytes may have been successful at the transition from an aquatic habitat to land, but they are still dependent on water for reproduction, and absorb moisture and nutrients through the gametophyte surface. The lack of roots for absorbing water and minerals from the soil, as well as a lack of reinforced conducting cells, limits bryophytes to small sizes. Although they may survive in reasonably dry conditions, they cannot reproduce and expand their habitat range in the absence of water. Vascular plants, on the other hand, can achieve enormous heights, thus competing successfully for light. Photosynthetic

organs become leaves, and pipe-like cells or vascular tissues transport water, minerals, and fixed carbon throughout the organism.

In seedless vascular plants, the diploid sporophyte is the dominant phase of the lifecycle. The gametophyte is now an inconspicuous, but still independent, organism. Throughout plant evolution, there is an evident reversal of roles in the dominant phase of the lifecycle. Seedless vascular plants still depend on water during fertilization, as the sperm must swim on a layer of moisture to reach the egg. This step in reproduction explains why ferns and their relatives are more abundant in damp environments.

Characteristics of Seedless Vascular Plants

As the name implies, this group of plants lack seeds and fruit but do have vascular tissue. The vascular tissue xylem and phloem are used to move water throughout the plant and there is lignin found in the cell walls for structural support and for enhanced water transport efficiency. However, water is still necessary for sperm to swim to the egg and why many ferns and their relatives are found in damp environments. Lastly, the sporophyte is the dominant phase of the life cycle with the gametophyte being and inconspicuous and independent organism.

Vascular Tissue: Xylem and Phloem

The first fossils that show the presence of vascular tissue date to the Silurian period, about 430 million years ago. The simplest arrangement of conductive cells shows a pattern of xylem at the center surrounded by phloem. **Xylem** is the tissue responsible for the storage and long-distance transport of water and nutrients, as well as the transfer of water-soluble growth factors from the organs of synthesis to the target organs. The tissue consists of conducting cells, known as tracheids, and supportive filler tissue, called parenchyma. Xylem conductive cells incorporate the compound **lignin** into their walls, and are thus described as lignified. Lignin itself is a complex polymer that is impermeable to water and strengthens vascular tissue. With their rigid cell walls, the xylem cells provide support to the plant and allow it to achieve impressive heights. Tall plants have a selective advantage by being able to reach unfiltered sunlight and disperse their spores or seeds further away, thus expanding their range. By growing higher than other plants, tall trees cast their shadow on shorter plants and limit competition for water and precious nutrients in the soil.

Phloem is the second type of vascular tissue; it transports sugars, proteins, and other solutes throughout the plant. Phloem cells are divided into sieve elements (conducting cells) and cells that support the sieve elements. Together, xylem and phloem tissues form the vascular system of plants.

Roots: Support for the Plant

Roots are not well preserved in the fossil record. Nevertheless, it seems that roots appeared later in evolution than vascular tissue. The development of an extensive network of roots represented a significant new feature of vascular plants. Thin rhizoids attached bryophytes to the substrate, but these rather flimsy filaments did not provide a strong anchor for the plant; neither did they absorb substantial amounts of water and nutrients. In contrast, roots, with their prominent vascular tissue system, transfer water and minerals from the soil to the rest of the plant. The extensive network of roots that penetrates deep into the soil to reach sources of water also stabilizes plants by acting as a ballast or anchor. The majority of roots establish a mulutalistic relationship with fungi, forming mycorrhizae, which benefit the plant by greatly increasing the surface area for absorption of water and soil minerals and nutrients.

Leaves, Sporophylls, and Strobili

A third adaptation marks the seedless vascular plants. Accompanying the prominence of the sporophyte and the development of vascular tissue, the evolution of true leaves improved photosynthetic efficiency. Leaves capture more sunlight with their increased surface area by employing more chloroplasts to trap light energy and convert it to chemical energy, which is then used to fix atmospheric carbon dioxide into carbohydrates. The carbohydrates are exported to the rest of the plant by the conductive cells of phloem tissue.

The existence of two types of morphology suggests that leaves evolved independently in several groups of plants. The first type of leaf is the microphyll, or “little leaf,” which can be dated to 350 million years ago in the late Silurian. A microphyll is small and has a simple vascular system. A single unbranched **vein**—a bundle of vascular tissue made of xylem and phloem—runs through the center of the leaf. Microphylls may have originated from the flattening of lateral branches, or from sporangia that lost their reproductive capabilities. Microphylls are present in the club mosses and probably preceded the development of megaphylls, or “big leaves”, which are larger leaves with a pattern of branching veins. Megaphylls most likely appeared independently several times during the course of evolution. Their complex networks of veins suggest that several branches may have combined into a flattened organ, with the gaps between the branches being filled with photosynthetic tissue.

In addition to photosynthesis, leaves play another role in the lives of plants. Pine cones, mature fronds of ferns, and flowers

are all sporophylls—leaves that were modified structurally to bear sporangia. Strobili are cone-like structures that contain sporangia. They are prominent in conifers and are commonly known as pine cones.

Ferns and Other Seedless Vascular Plants

By the late Devonian period, plants had evolved vascular tissue, well-defined leaves, and root systems. With these advantages, plants increased in height and size. During the Carboniferous period, swamp forests of club mosses and horsetails—some specimens reaching heights of more than 30 m (100 ft)—covered most of the land. These forests gave rise to the extensive coal deposits that gave the Carboniferous its name. In seedless vascular plants, the sporophyte became the dominant phase of the lifecycle.

Water is still required for fertilization of seedless vascular plants, and most favor a moist environment. Modern-day seedless vascular plants include club mosses, horsetails, ferns, and whisk ferns.

Phylum Lycopodiophyta: Club Mosses

The **club mosses**, or phylum Lycopodiophyta, are the earliest group of seedless vascular plants. They dominated the landscape of the Carboniferous, growing into tall trees and forming large swamp forests. Today's club mosses are diminutive, evergreen plants consisting of a stem (which may be branched) and microphylls (**Figure 21.17**). The phylum Lycopodiophyta consists of close to 1,200 species, including quillworts (*Isoetales*), club mosses (*Lycopodiales*), and spike mosses (*Selaginellales*), which, despite their common names, are not true mosses (bryophytes).

Lycophytes follow the pattern of alternation of generations seen in the bryophytes, except that the sporophyte is the major stage of the lifecycle. The gametophytes do not depend on the sporophyte for nutrients. Some gametophytes develop underground and form mycorrhizal associations with fungi. In club mosses, the sporophyte gives rise to sporophylls arranged in strobili, cone-like structures that give the class its name. Lycophytes can be homosporous or heterosporous.

Figure 21.17 In the club mosses such as *Lycopodium clavatum*, sporangia are arranged in clusters called strobili. (credit: Cory Zanker)

Phylum Monilophyta: Class Equisetopsida (Horsetails)

Horsetails, whisk ferns and ferns belong to the phylum Monilophyta, with **horsetails** placed in the Class Equisetopsida. The modern genus *Equisetum* contains the only survivors of a large and diverse group of horsetails known as the Arthrophyta, which produced large trees and entire swamp forests during the Carboniferous. The plants are usually found in damp environments and marshes (**Figure 21.18**).

Figure 21.18 Horsetails thrive in a marsh. (credit: Myriam Feldman)

The stem of a horsetail is characterized by the presence of joints or nodes, hence the name Arthrophyta (arthro- = "joint"; -phyta = "plant"). Leaves and branches come out as whorls from the evenly spaced joints. The needle-shaped leaves do not contribute greatly to photosynthesis, the majority of which takes place in the green stem (**Figure 21.19**).

Figure 21.19 Thin leaves originating at the joints are noticeable on the horsetail plant. Horsetails were once used as scrubbing brushes and were nicknamed scouring brushes. (credit: Myriam Feldman)

Silica collects in the epidermal cells, contributing to the stiffness of horsetail plants. Underground stems known as rhizomes anchor the plants to the ground. Modern-day horsetails are homosporous and an individual produces both male and female gametes.

Phylum Monilophyta: Class Psilotopsida (Whisk Ferns)

While most ferns form large leaves and branching roots, the **whisk ferns**, Class Psilotopsida, lack both roots and leaves. Photosynthesis takes place in their green stems, and small yellow knobs form at the tip of the branch stem and contain the sporangia.

Figure 21.20 The whisk fern *Psilotum nudum* has conspicuous green stems with knob-shaped sporangia. (credit: Forest & Kim Starr)

Phylum Monilophyta: Class Psilotopsida (Ferns)

Ferns are the most readily recognizable seedless vascular plants and are considered the most advanced seedless vascular plants because they display characteristics commonly observed in seed plants. More than 20,000 species of ferns live in environments ranging from tropics to temperate forests. Although some species survive in dry environments, most ferns are restricted to moist, shaded places. Ferns made their appearance in the fossil record during the Devonian period and expanded during the Carboniferous.

The dominant stage of the lifecycle of a fern is the sporophyte, which consists of large compound leaves called fronds. Fronds fulfill a double role; they are photosynthetic organs that also carry reproductive organs. The stem may be buried underground as a rhizome, from which adventitious roots grow to absorb water and nutrients from the soil; or, they may grow above ground as a trunk in tree ferns (**Figure 21.21**). Adventitious organs are those that grow in unusual places, such as roots growing from the side of a stem.

Figure 21.21 Some specimens of this short tree-fern species can grow very tall. (credit: Adrian Pingstone)

The tip of a developing fern frond is rolled into a crozier, or fiddlehead (**Figure 21.22a** and **Figure 21.22b**). Fiddleheads unroll as the frond develops.

(a) (b)

Figure 21.22 Croziers, or fiddleheads, are the tips of fern fronds. (credit a: modification of work by Cory Zanker; credit b: modification of work by Myriam Feldman)

The lifecycle of a fern is depicted in **Figure 21.23**.

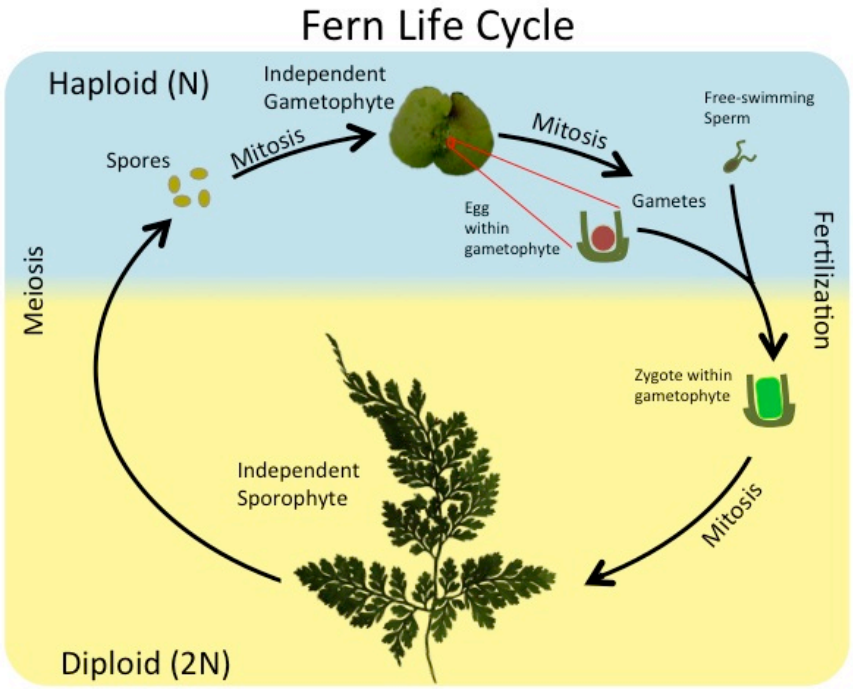


Figure 21.23 This life cycle of a fern shows alternation of generations with a dominant sporophyte stage. Work by Eva Horne and Robert A. Bear

Most ferns produce a single type of spore and are therefore homosporous. The diploid sporophyte is the most conspicuous stage of the lifecycle. On the underside of its mature fronds, sori (singular, sorus) form as small clusters where sporangia develop (**Figure 21.24**).

Figure 21.24 Sori appear as small bumps on the underside of a fern frond. (credit: Myriam Feldman)

Inside the sori, spores are produced by meiosis and released into the air. Those that land on a suitable substrate germinate and form a heart-shaped gametophyte, which is attached to the ground by thin filamentous rhizoids (**Figure 21.25**).

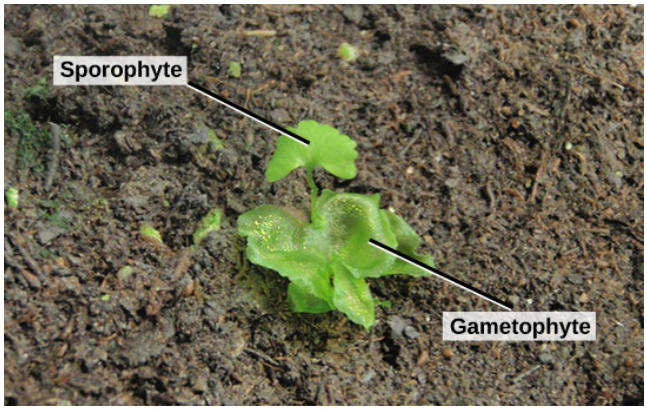


Figure 21.25 Shown here are a young sporophyte (upper part of image) and a heart-shaped gametophyte (bottom part of image). (credit: modification of work by "Vlmastra"/Wikimedia Commons)

The inconspicuous gametophyte harbors both sex gametangia. Flagellated sperm released from the antheridium swim on a wet surface to the archegonium, where the egg is fertilized. The newly formed zygote grows into the next generation sporophyte.

The Importance of Bryophytes and Seedless Vascular Plants

Mosses and liverworts are often the first macroscopic organisms to colonize an area, both in a primary succession—where bare land is settled for the first time by living organisms—or in a secondary succession, where soil remains intact after a catastrophic event wipes out many existing species. Their spores are carried by the wind, birds, or insects. Once mosses and liverworts are established, they provide food and shelter for other species. In a hostile environment, like the tundra where the soil is frozen, bryophytes grow well because they do not have roots and can dry and rehydrate rapidly once water is again available. Mosses are at the base of the food chain in the tundra biome. Many species—from small insects to musk oxen and reindeer—depend on mosses for food. In turn, predators feed on the herbivores, which are the primary consumers. Some reports indicate that bryophytes make the soil more amenable to colonization by other plants. Because they establish symbiotic relationships with nitrogen-fixing cyanobacteria, mosses replenish the soil with nitrogen.

At the end of the nineteenth century, scientists observed that lichens and mosses were becoming increasingly rare in urban and suburban areas. Since bryophytes have neither a root system for absorption of water and nutrients, nor a cuticle layer that protects them from desiccation, pollutants in rainwater readily penetrate their tissues; they absorb moisture and nutrients through their entire exposed surfaces. Therefore, pollutants dissolved in rainwater penetrate plant tissues readily and have a larger impact on mosses than on other plants. The disappearance of mosses can be considered a bioindicator for the level of pollution in the environment.

Ferns contribute to the environment by promoting the weathering of rock, accelerating the formation of topsoil, and slowing down erosion by spreading rhizomes in the soil. The water ferns of the genus *Azolla* harbor nitrogen-fixing cyanobacteria and restore this important nutrient to aquatic habitats.

Seedless plants have historically played a role in human life through use as tools, fuel, and medicine. Dried peat moss,

Sphagnum, is commonly used as fuel in some parts of Europe and is considered a renewable resource. *Sphagnum* bogs (**Figure 21.26**) are cultivated with cranberry and blueberry bushes. The ability of *Sphagnum* to hold moisture makes the moss a common soil conditioner. Florists use blocks of *Sphagnum* to maintain moisture for floral arrangements.

Figure 21.26 *Sphagnum acutifolium* is dried peat moss and can be used as fuel. (credit: Ken Goulding)

The attractive fronds of ferns make them a favorite ornamental plant. Because they thrive in low light, they are well suited as house plants. More importantly, fiddleheads are a traditional spring food of Native Americans in the Pacific Northwest, and are popular as a side dish in French cuisine. The licorice fern, *Polypodium glycyrrhiza*, is part of the diet of the Pacific Northwest coastal tribes, owing in part to the sweetness of its rhizomes. It has a faint licorice taste and serves as a sweetener. The rhizome is also valued by Native Americans for its medicinal properties and is used as a remedy for sore throat.

By far the greatest impact of seedless vascular plants on human life, however, comes from their extinct progenitors. The tall club mosses, horsetails, and tree-like ferns that flourished in the swampy forests of the Carboniferous period gave rise to large deposits of coal throughout the world. Coal provided an abundant source of energy during the Industrial Revolution, which had tremendous consequences on human societies, including rapid technological progress and growth of large cities, as well as the degradation of the environment. Coal is still a prime source of energy and also a major contributor to global warming.

21.5 | Evolution of Seed Plants

Introduction

> “ I was aware of Darwin's views fourteen years before I adopted them and I have done so solely and entirely from an independent study of the plants themselves. ”

Sir Joseph Dalton Hooker, letter to E.H. Harvey, in L. Huxley, *Life and Letters of Sir Joseph Dalton Hooker* (1918)

The evolution of plants is indeed well-understood, and Hooker's recognition of the usefulness of evolutionary theory in understanding plant biology was correct. The first plants to colonize land were most likely closely related to modern day mosses (bryophytes) and are thought to have appeared about 500 million years ago, based on fossil and DNA sequence evidence. They were followed by liverworts (also bryophytes) and primitive vascular plants—the pterophytes—from which modern ferns are derived. Like gymnosperms and angiosperms, the life cycle of bryophytes and pterophytes is characterized by the alternation of generations. What sets bryophytes and pterophytes apart from gymnosperms and angiosperms is their reproductive requirement for water. The completion of the bryophyte and pterophyte life cycle requires water because the male gametophyte releases sperm, which must swim—propelled by their flagella—to reach and fertilize the female gamete or egg. After fertilization, the zygote matures and grows into a sporophyte, which in turn will form sporangia or "spore vessels." In the sporangia, mother cells undergo meiosis and produce the haploid spores. Release of spores in a suitable environment will lead to germination and a new generation of gametophytes.

In seed plants, the evolutionary trend led to a dominant sporophyte generation, and at the same time, a systematic reduction

in the size of the gametophyte: from a conspicuous structure to a microscopic cluster of cells enclosed in the tissues of the sporophyte. The gametophyte is thus dependent on the sporophyte for shelter and nutrition. Whereas lower vascular plants, such as club mosses and ferns, are mostly homosporous (produce only one type of spore), all seed plants are heterosporous. They form two types of spores: megaspores (female) and microspores (male). Megaspores develop into female gametophytes that produce eggs, and microspores mature into male gametophytes that generate sperm. Because the gametophytes mature within the sporophyte, they are not free-living, as are the gametophytes of other seedless vascular plants. Heterosporous seedless plants are seen as the evolutionary forerunners of seed plants.

Seeds and pollen—two critical adaptations to drought, and to reproduction that doesn't require water—distinguish seed plants from other (seedless) vascular plants. Both adaptations were required for the colonization of land begun by the bryophytes and their ancestors. Fossils place the earliest distinct seed plants at about 350 million years ago. The first reliable record of gymnosperms dates their appearance to the late Paleozoic, about 319 million years ago (**Figure 21.27**). Gymnosperms were preceded by progymnosperms, the first naked seed plants, which arose about 380 million years ago. Progymnosperms were a transitional group of plants that superficially resembled conifers (cone bearers) because they produced wood from the secondary growth of vascular tissues; however, they still reproduced like ferns, releasing spores into the environment. Gymnosperms dominated the landscape in the early and middle Mesozoic era. Angiosperms surpassed gymnosperms by about 100 million years ago in the late Mesozoic era, and today are the most abundant plant group in most terrestrial biomes.

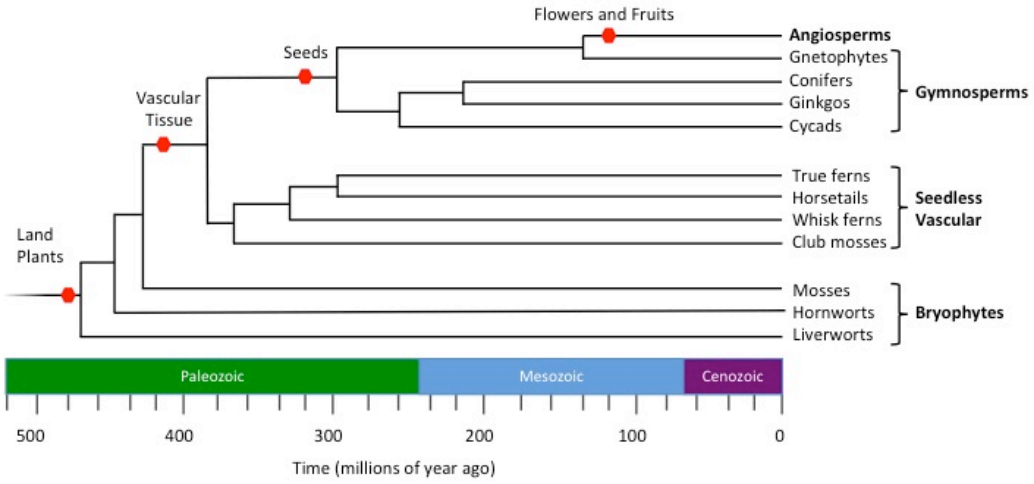


Figure 21.27 Phylogeny of the land plants with the major adaptation for living on land. Note the geological eras along the bottom. Work by Robert A. Bear

Pollen and seed were adaptations that allowed seed plants to break their dependence on water for reproduction and development of the embryo, and to conquer dry land. The **pollen grains** are the male gametophytes, which contain the sperm (gametes) of the plant. The small haploid (1N) cells are encased in a protective coat that prevents desiccation (drying out) and mechanical damage. Pollen grains can travel far from their original sporophyte, spreading the plant's genes. The **seed** offers the embryo protection, nourishment, and a mechanism to maintain dormancy for tens or even thousands of years, ensuring germination can occur when growth conditions are optimal. Seeds therefore allow plants to disperse the next generation through both space and time. With such evolutionary advantages, seed plants have become the most successful and familiar group of plants, in part because of their size and striking appearance.

Evolution of Gymnosperms

The fossil plant *Elkinsia polymorpha*, a "seed fern" from the Devonian period—about 400 million years ago—is considered the earliest seed plant known to date. Seed ferns (**Figure 21.28**) produced seeds along their branches without specialized structures. What makes them the first true seed plants is that they developed structures to enclose and protect the **ovule**—the female gametophyte and associated tissues—which develops into a seed upon fertilization. Seed plants resembling modern tree ferns became more numerous and diverse in the coal swamps of the Carboniferous period.

Figure 21.28 This fossilized leaf is from *Glossopteris*, a seed fern that thrived during the Permian age (290–240 million years ago). (credit: D.L. Schmidt, USGS)

Fossil records indicate the first gymnosperms (progymnosperms) most likely originated in the Paleozoic era, during the middle Devonian period: about 390 million years ago. Following the wet Mississippian and Pennsylvanian periods, which were dominated by giant fern trees, the Permian period was dry. This gave a reproductive edge to seed plants, which are better adapted to survive dry spells. The Ginkgoales, a group of gymnosperms with only one surviving species— *Gingko biloba*—were the first gymnosperms to appear during the lower Jurassic. Gymnosperms expanded in the Mesozoic era (about 240 million years ago), supplanting ferns in the landscape, and reaching their greatest diversity during this time. The Jurassic period was as much the age of the cycads (palm-tree-like gymnosperms) as the age of the dinosaurs. Gingkoales and more familiar conifers also dotted the landscape. Although angiosperms (flowering plants) are the major form of plant life in most biomes, gymnosperms still dominate some ecosystems, such as the taiga (boreal forests) and the alpine forests at higher mountain elevations (**Figure 21.29**) because of their adaptation to cold and dry growth conditions.

Figure 21.29 This boreal forest (taiga) has low-lying plants and conifer trees. (credit: L.B. Brubaker, NOAA)

Seeds and Pollen as an Evolutionary Adaptation to Dry Land

Unlike bryophyte and fern spores (which are haploid cells dependent on moisture for rapid development of gametophytes), seeds contain a diploid embryo that will germinate into a sporophyte. Storage tissue to sustain growth and a protective coat give seeds their superior evolutionary advantage. Several layers of hardened tissue prevent desiccation, and free reproduction from the need for a constant supply of water. Furthermore, seeds remain in a state of dormancy—induced by desiccation and the hormone abscisic acid—until conditions for growth become favorable. Whether blown by the wind, floating on water, or carried away by animals, seeds are scattered in an expanding geographic range, thus avoiding competition with the parent plant.

Pollen grains (**Figure 21.30**) are male gametophytes and are carried by wind, water, or a pollinator. The whole structure is protected from desiccation and can reach the female organs without dependence on water. Male gametes reach female gametophyte and the egg cell though a pollen tube: an extension of a cell within the pollen grain. The sperm of modern gymnosperms lack flagella, but in cycads and the *Gingko*, the sperm still possess flagella that allow them to swim down the **pollen tube** to the female gamete; however, they are enclosed in a pollen grain.

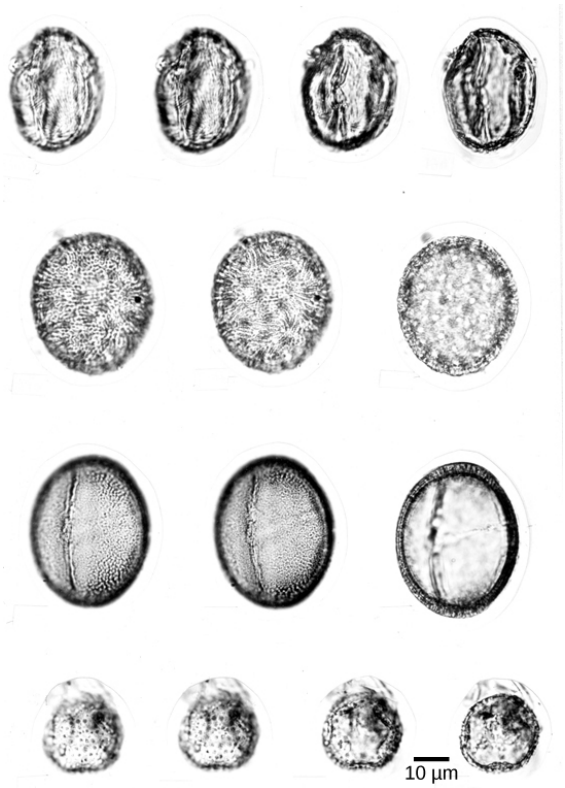


Figure 21.30 This fossilized pollen is from a Buckbean fen core found in Yellowstone National Park, Wyoming. The pollen is magnified 1,054 times. (credit: R.G. Baker, USGS; scale-bar data from Matt Russell)

Evolution of Angiosperms

Undisputed fossil records place the massive appearance and diversification of angiosperms in the middle to late Mesozoic era. Angiosperms ("seed in a vessel") produce a flower containing male and/or female reproductive structures. Fossil evidence (**Figure 21.31**) indicates that flowering plants first appeared in the Lower Cretaceous, about 125 million years ago, and were rapidly diversifying by the Middle Cretaceous, about 100 million years ago. Earlier traces of angiosperms are scarce. Fossilized pollen recovered from Jurassic geological material has been attributed to angiosperms. A few early Cretaceous rocks show clear imprints of leaves resembling angiosperm leaves. By the mid-Cretaceous, a staggering number of diverse flowering plants crowd the fossil record. The same geological period is also marked by the appearance of many modern groups of insects, including pollinating insects that played a key role in ecology and the evolution of flowering plants.

Although several hypotheses have been offered to explain this sudden profusion and variety of flowering plants, none have garnered the consensus of paleobotanists (scientists who study ancient plants). New data in comparative genomics and paleobotany have, however, shed some light on the evolution of angiosperms. The two adaptations of flowers and fruit represent an improved reproductive strategy that served to protect the embryo, while increasing genetic variability and range. Angiosperms did not evolve from gymnosperms, but the groups do share a common ancestor (**Figure 21.27**).

The most primitive living angiosperm is considered to be *Amborella trichopoda*, a small plant native to the rainforest of New Caledonia, an island in the South Pacific. Analysis of the genome of *A. trichopoda* has shown that it is related to all existing flowering plants and belongs to the oldest confirmed branch of the angiosperm family tree. A few other angiosperm groups called basal angiosperms, are viewed as primitive because they branched off early from the phylogenetic tree. Most modern angiosperms are classified as either monocots or eudicots, based on the structure of their leaves and embryos. Basal angiosperms, such as water lilies, are considered more primitive because they share morphological traits with both monocots and eudicots.

Figure 21.31 This leaf imprint shows a *Ficus speciosissima*, an angiosperm that flourished during the Cretaceous period. (credit: W. T. Lee, USGS)

Flowers and Fruits as an Evolutionary Adaptation

Angiosperms produce their gametes in separate organs, which are usually housed in a **flower**. Both fertilization and embryo development take place inside an anatomical structure that provides a stable system of sexual reproduction largely sheltered from environmental fluctuations. Flowering plants are the most diverse phylum on Earth after insects (Arthropoda); flowers come in a bewildering array of sizes, shapes, colors, smells, and arrangements. Most flowers have a mutualistic pollinator, with the distinctive features of flowers reflecting the nature of the pollination agent. The relationship between pollinator and flower characteristics is one of the great examples of coevolution.

Following fertilization of the egg, the ovule grows into a seed. The surrounding tissues of the ovary thicken, developing into a **fruit** that will protect the seed and often ensure its dispersal over a wide geographic range. Not all fruits develop from an ovary; such structures are “false fruits.” Like flowers, fruit can vary tremendously in appearance, size, smell, and taste. Tomatoes, walnut shells and avocados are all examples of fruit. As with pollen and seeds, fruits also act as agents of dispersal. Some may be carried away by the wind. Many attract animals that will eat the fruit and pass the seeds through their digestive systems, then deposit the seeds in another location. Cockleburs are covered with stiff, hooked spines that can hook into fur (or clothing) and hitch a ride on an animal for long distances. The cockleburs that clung to the velvet trousers of an enterprising Swiss hiker, George de Mestral, inspired his invention of the loop and hook fastener he named Velcro.

21.6 | Gymnosperms

Introduction

> “ It took more than three thousand years to make some of the trees in these western woods ... Through all the wonderful, eventful centuries since Christ's time—and long before that—God has cared for these trees, saved them from drought, disease, avalanches, and a thousand straining, leveling tempests and floods; but he cannot save them from fools.”

John Muir, in "The American Forests", *Atlantic Monthly* (Aug 1897)

Figure 21.32 Giant Sequoia The Grizzly Giant, a Giant Sequoia *(Sequoiadendron giganteum)* in the Mariposa Grove of Yosemite National Park. John Muir and others petitioned the US Congress, resulting in the creation of Yosemite and Sequoia National Parks in 1890. Photo by David A. Rintoul.

The redwood trees and Douglas firs referred to in Muir's quote are indeed ancient, and among the largest living things on the planet. They are members of the group we call **gymnosperms**, meaning “naked seeds,” which are a diverse group of seed plants and are **paraphyletic**. Paraphyletic groups are those in which not all members are descendants of a single common ancestor. Gymnosperm characteristics include naked seeds (not enclosed in an ovary), separate female and male gametes, pollination by wind, and tracheids (which transport water and solutes in the vascular system).

Gymnosperm seeds are not enclosed in an ovary (they are "naked"); rather, they are exposed on cones or modified leaves. Gymnosperms were the dominant phylum in the Mesozoic era. They are adapted to live where fresh water is scarce during part of the year, or in the nitrogen-poor soil of a bog. Therefore, they are still the prominent phylum in the coniferous biome or taiga, where the evergreen conifers have a selective advantage in cold and dry weather. Evergreen conifers continue low levels of photosynthesis during the cold months, and are ready to take advantage of the first sunny days of spring. One disadvantage is that conifers are more susceptible than deciduous trees to infestations because conifers do not lose their leaves all at once. They cannot, therefore, shed parasites and restart with a fresh supply of leaves in spring.

The life cycle of a gymnosperm involves alternation of generations, with a dominant sporophyte and reduced gametophytes that resides within the sporophyte. All gymnosperms are heterosporous. The male and female reproductive organs can form in cones (strobili). Male and female sporangia are produced either on the same plant, described as monoecious (“one home”), or on separate plants, referred to as dioecious (“two homes”) plants. The life cycle of a conifer will serve as our example of reproduction in gymnosperms.

Life Cycle of a Conifer

Pine trees are conifers (cone bearing) and carry both male and female sporophylls on the same mature sporophyte.

Therefore, they are monoecious plants. Like all gymnosperms, pines are heterosporous and generate two different types of spores: male microspores and female megaspores. In the male or staminate cones, the microsporocytes give rise to pollen grains by meiosis. In the spring, large amounts of yellow pollen are released and carried by the wind. Some gametophytes will land on a female cone. Pollination is defined as the initiation of pollen tube growth. The pollen tube develops slowly, and the generative cell in the pollen grain divides into two haploid sperm cells by mitosis. At fertilization, one of the sperm cells will finally unite its haploid nucleus with the haploid nucleus of a haploid egg cell.

Female cones contain two ovules per scale. One megaspore mother cell, or megasporocyte, undergoes meiosis in each ovule. Three of the four cells break down; only a single surviving cell will develop into a female multicellular gametophyte, which encloses one or more archegonia (an archegonium is a reproductive organ that contains a single large egg). Upon fertilization, the resulting diploid zygote will give rise to the embryo, which is enclosed in a seed coat of tissue from the parent plant. Fertilization and seed development is a long process in pine trees: it may take up to two years after pollination. The seed that is formed contains three generations of tissues: the seed coat that originates from the sporophyte tissue, the gametophyte that will provide nutrients, and the embryo itself.

Figure 21.33 illustrates the life cycle of a conifer. The sporophyte (2N) phase is the longest phase in the life of a gymnosperm. The gametophytes (1N)—microspores and megaspores—are reduced in size. It may take more than year between pollination and fertilization while the pollen tube grows towards the megasporocyte (2N), which undergoes meiosis into megaspores. The megaspores will mature into female gametophytes that then produce eggs (1N).

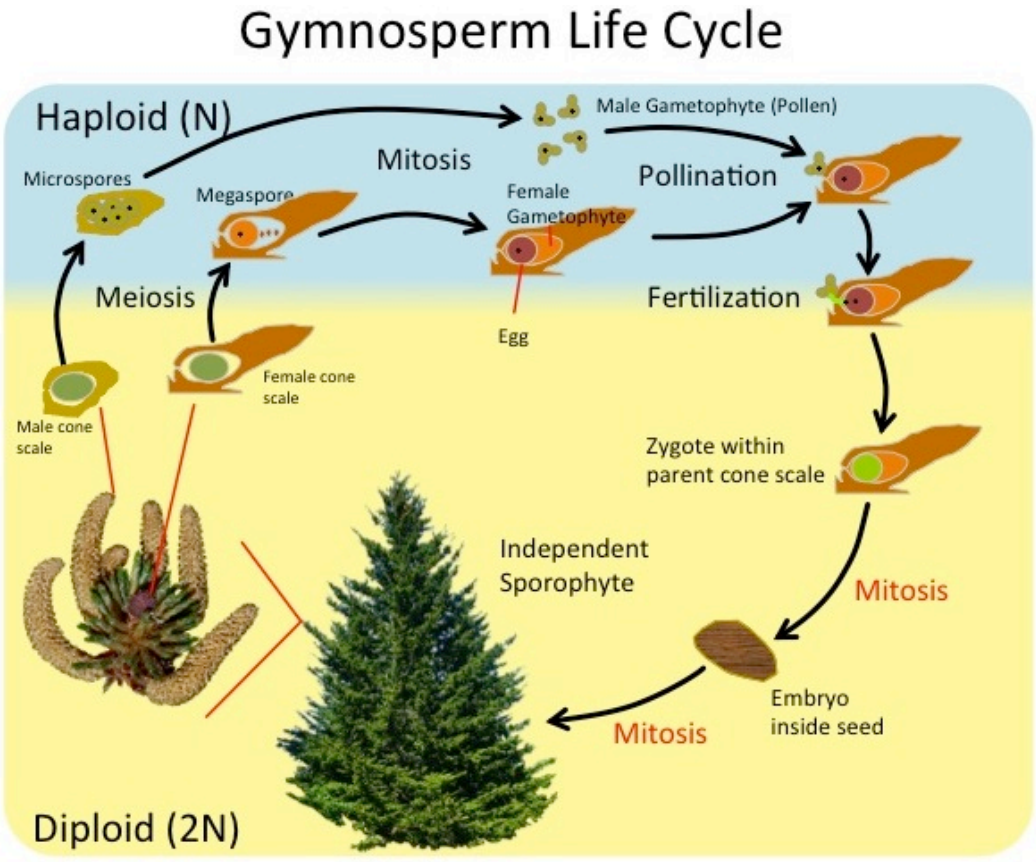


Figure 21.33 This image shows the life cycle of a conifer. Pollen from male cones blows up into upper branches, where it fertilizes female cones. Work by Eva Horne and Robert A. Bear

Diversity of Gymnosperms

Modern gymnosperms are classified into four phyla. Coniferophyta, Cycadophyta, and Ginkgophyta are similar in their production of secondary cambium (cells that generate the vascular system of the trunk or stem and are partially specialized for water transportation) and their pattern of seed development. However, the three phyla are not closely related phylogenetically to each other. Gnetophyta are considered the closest group to angiosperms because they produce true xylem tissue.

Conifers (Coniferophyta)

Conifers are the dominant phylum of gymnosperms, with the most variety of species (**Figure 21.34**). Most are typically tall trees that usually bear scale-like or needle-like leaves. Water evaporation from leaves is reduced by their thin shape and the thick cuticle. Snow slides easily off needle-shaped leaves, keeping the load light and decreasing breaking of branches. Adaptations to cold and dry weather explain the predominance of conifers at high altitudes and in cold climates. Conifers include familiar evergreen trees such as pines, spruces, firs, cedars, sequoias, and yews. A few species are deciduous and lose their leaves in fall. The European larch and the tamarack are examples of deciduous conifers (**Figure 21.34d**). Many coniferous trees are harvested for paper pulp and timber. The wood of conifers is more primitive than the wood of angiosperms; it contains tracheids, but no vessel elements, and is therefore referred to as "soft wood."

Figure 21.34 Conifers are the dominant form of vegetation in cold or arid environments and at high altitudes. Shown here are the (a) evergreen spruce *Picea* sp., (b) juniper *Juniperus* sp., (c) coast redwood *Sequoia sempervirens*, which is a deciduous gymnosperm, and (d) the tamarack *Larix larcinia*. Notice the yellow leaves of the tamarack. (credit a: modification of work by Rosendahl; credit b: modification of work by Alan Levine; credit c: modification of work by Wendy McCormic; credit d: modification of work by Micky Zlimen)

Cycads

Cycads thrive in mild climates, and are often mistaken for palms because of the shape of their large, compound leaves. Cycads bear large cones (**Figure 21.35**), and may be pollinated by beetles rather than wind: unusual for a gymnosperm. They dominated the landscape during the age of dinosaurs in the Mesozoic, but only a hundred or so species persisted to modern times. They face possible extinction, and several species are protected through international conventions. Because of their attractive shape, they are often used as ornamental plants in gardens in the tropics and subtropics.

Figure 21.35 This *Encephalartos ferox* cycad has large cones and broad, fern-like leaves. (credit: Wendy Cutler)

Gingkophytes

The single surviving species of **Gingkophytes** is the *Gingko biloba* (**Figure 21.36**). Its fan-shaped leaves—unique among seed plants because they feature a dichotomous venation pattern—turn yellow in autumn and fall from the tree. For centuries, *G. biloba* was cultivated by Chinese Buddhist monks in monasteries, which ensured its preservation. It is planted in public spaces because it is unusually resistant to pollution. Male and female organs are produced on separate plants. Typically, gardeners plant only male trees because the seeds produced by the female plant have an off-putting smell of rancid butter.

Figure 21.36 This plate from the 1870 book *Flora Japonica, Sectio Prima (Tafelband)* depicts the leaves and fruit of *Gingko biloba*, as drawn by Philipp Franz von Siebold and Joseph Gerhard Zuccarini.

Gnetophytes

Gnetophytes are the closest relative to modern angiosperms, and include three dissimilar genera of plants: *Ephedra*, *Gnetum*, and *Welwitschia* (**Figure 21.37**). Like angiosperms, they have broad leaves. In tropical and subtropical zones, gnetophytes are vines or small shrubs. *Ephedra* occurs in dry areas of the West Coast of the United States and Mexico. *Ephedra*'s small, scale-like leaves are the source of the compound ephedrine, which is used in medicine as a potent

decongestant. Because ephedrine is similar to amphetamines, both in chemical structure and neurological effects, its use is restricted to prescription drugs. Like angiosperms, but unlike other gymnosperms, all gnetophytes possess vessel elements in their xylem.

(a) *Ephedra* (b) *Gnetum* (c) *Welwitschia*

Figure 21.37 (a) *Ephedra viridis*, known by the common name Mormon tea, grows on the West Coast of the United States and Mexico. (b) *Gnetum gnemon* grows in Malaysia. (c) The large *Welwitschia mirabilis* can be found in the Namibian desert. (credit a: modification of work by USDA; credit b: modification of work by Malcolm Manners; credit c: modification of work by Derek Keats)

21.7 | Angiosperms

Introduction

> "The ginkgo tree is from the era of dinosaurs, but while the dinosaur has been extinguished, the modern ginkgo has not changed. After the atomic bomb in Hiroshima, the ginkgo was the first tree that came up. It's amazing."

Koji Nakanishi, organic chemist (2013)

From their humble and still obscure beginning during the early Jurassic period, the angiosperms—or flowering plants—have evolved to dominate most terrestrial ecosystems (**Figure 21.38**). With more than 250,000 species, the angiosperm phylum (Anthophyta) is second only to insects in terms of diversification.

Figure 21.38 These flowers grow in a botanical garden border in Bellevue, WA. Flowering plants dominate terrestrial landscapes. The vivid colors of flowers are an adaptation to pollination by animals such as insects and birds. (credit: Myriam Feldman)

The success of angiosperms is due to two novel reproductive structures: flowers and fruit. The function of the flower is to ensure pollination. Flowers also provide protection for the ovule and developing embryo inside a receptacle. The function of the fruit is seed dispersal. They also protect the developing seed. Different fruit structures or tissues on fruit—such as sweet flesh, wings, parachutes, or spines that grab—reflect the dispersal strategies that help spread seeds.

Flowers

Flowers are modified leaves, or sporophylls, organized around a central stalk. Although they vary greatly in appearance, all flowers contain the same structures: sepals, petals, carpels, and stamens. The receptacle attaches the flower to the plant. A whorl of **sepals** (collectively called the **calyx**) is located at the base of the peduncle (stem of the flower) and encloses the unopened floral bud. Sepals are usually photosynthetic organs, although there are some exceptions. For example, the corolla in lilies and tulips consists of three sepals and three petals that look virtually identical. **Petals**, collectively the **corolla**, are located inside the whorl of sepals and often display vivid colors to attract pollinators. Flowers pollinated by wind are usually small, feathery, and visually inconspicuous. The sexual organs (carpels and stamens) are located at the center of the flower.

As illustrated in **Figure 21.39**, styles, stigmas, and ovules constitute the female organ: the **carpel**. Flower structure is very diverse, and carpels may be singular, multiple, or fused. Multiple fused carpels comprise a **pistil**. The megaspores and the female gametophytes are produced and protected by the thick tissues of the carpel. A long, thin structure called a **style** leads from the sticky **stigma**, where pollen is deposited, to the **ovary**, enclosed in the carpel. The ovary houses one or more ovules, each of which will develop into a seed upon fertilization. The male reproductive organs, the **stamens**, surround the central carpel. Stamens are composed of a thin stalk called a **filament** and a sac-like structure called the anther. The filament supports the **anther**, where the microspores are produced by meiosis and develop into pollen grains.

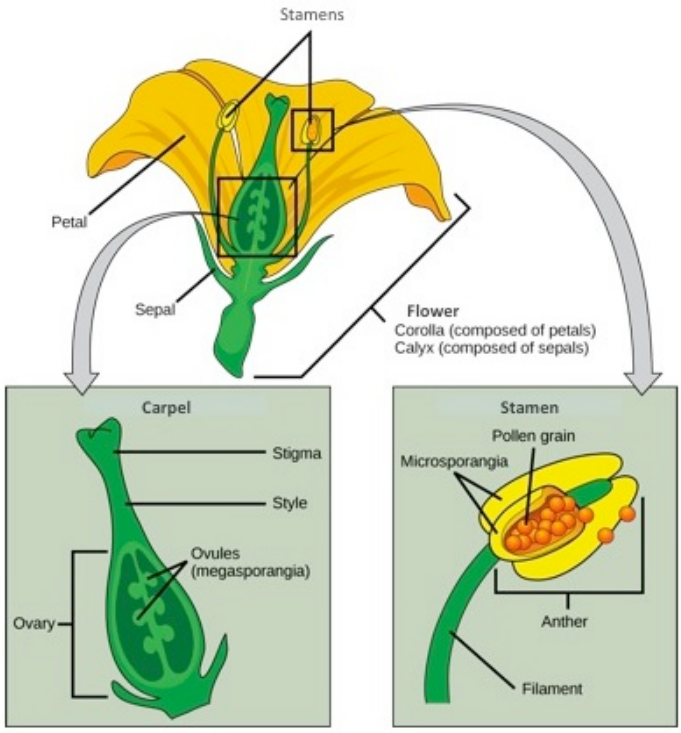


Figure 21.39 This image depicts the structure of a perfect flower. Perfect flowers produce both male and female floral organs. The flower shown has only one carpel, but some flowers have a cluster of carpels. Together, all the carpels make up the pistil. (credit: modification of work by Mariana Ruiz Villareal)

Fruit

As the seed develops, the walls of the ovary thicken and form the fruit. The seed forms in an ovary, which also enlarges as the seeds grow. In botany, a fertilized and fully grown, ripened ovary is a fruit. Many foods commonly called vegetables are actually fruit. Eggplants, zucchini, string beans, and bell peppers are all technically fruit because they contain seeds and are derived from the thick ovary tissue. Acorns (a nut), and winged maple whirligigs (whose botanical name is samara) are also fruit. Botanists classify fruit into more than two dozen different categories, only a few of which are actually fleshy and sweet.

Mature fruit can be fleshy or dry. Fleshy fruit include the familiar berries, peaches, apples, grapes, and tomatoes. Rice, wheat, and nuts are examples of dry fruit. Another distinction is that not all fruits are derived from the ovary. For instance, strawberries are derived from the receptacle and in apples the receptacle forms the pericarp or the fleshy part. Some fruits are derived from separate ovaries in a single flower, such as the raspberry. Other fruits, such as the pineapple, form from clusters of flowers. Additionally, some fruits, like watermelon and orange, have rinds. Regardless of how they are formed, fruits are an agent of seed dispersal. The variety of shapes and characteristics reflect the mode of dispersal. Wind carries the light dry fruit of trees and dandelions. Water transports floating coconuts. Some fruits attract herbivores with color or perfume, or as food. Once eaten, tough, undigested seeds are dispersed through the herbivore's feces. Other fruits have burs and hooks to cling to fur and hitch rides on animals.

The Life Cycle of an Angiosperm

The sporophyte phase is the main phase of an angiosperm's life cycle (**Figure 21.40**). Like gymnosperms, angiosperms are heterosporous. Therefore, they generate microspores, which will generate pollen grains as the male gametophytes, and megaspores, which will form an ovule that contains female gametophytes. Inside the anthers' microsporangia, male gametophytes divide by meiosis to generate haploid microspores, which, in turn, undergo mitosis and give rise to pollen grains. Each pollen grain contains two cells: one generative cell that will divide into two sperm and a second cell that will become the pollen tube cell.

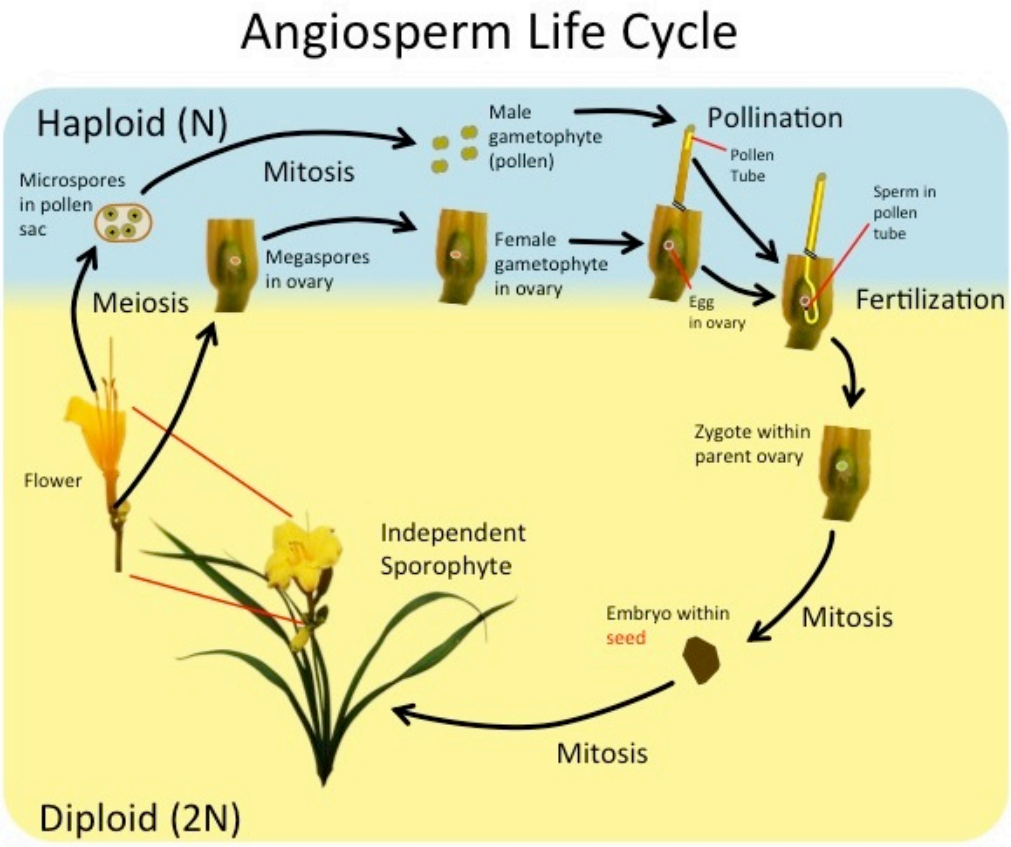


Figure 21.40 The life cycle of an angiosperm is shown. Anthers and carpels are structures that shelter the actual gametophytes: the pollen grain and embryo sac. Work by Eva Horne and Robert A. Bear

Within the ovule, sheltered within the ovary of the carpel, a cell undergoes meiosis, generating four megaspores—three small and one large. Only the large megaspore survives; it divides to produce the female gametophyte. The megaspore divides three times to form an eight-cell stage. Three of these cells migrate to each pole of the embryo sac while two come to the equator. These two cells will fuse to form one central cell with two haploid polar nuclei. The two cells closest to the egg are called synergids; the three cells on the opposite end of the gametophyte are called antipodals.

Pollen grains are the male gametophytes. When a pollen grain reaches the stigma, a pollen tube extends from the grain, grows down the style, and enters through the micropyle: an opening in the outer covering of the ovule. Two sperm cells travel down this tube and enter the ovule.

A double fertilization event then occurs. One sperm combines with the egg, forming a diploid zygote—the future embryo. The other sperm fuses with the two polar nuclei, forming a triploid (3n) cell that will develop into the endosperm, which is tissue that serves as a food reserve. The zygote develops into an embryo with a radicle, or small root, and one (monocot) or two (dicot) leaf-like organs called cotyledons. This difference in the number of embryonic leaves is the basis for the two major groups of angiosperms: the monocots and the eudicots. Seed food reserves are stored outside the embryo, in the form of complex carbohydrates, lipids, or proteins. The cotyledons serve as conduits to transmit the broken-down food reserves from their storage site inside the seed to the developing embryo. The seed consists of a toughened layer of integuments (the seed coat), the endosperm with food reserves, and at the center, the well-protected embryo.

Most flowers are monoecious, which means that they carry both stamens and carpels. Monoecious flowers are also known as “perfect” flowers because they contain both types of sex organs. However, only a few species self-pollinate. Both anatomical and environmental barriers promote pollination between different individuals (cross-pollination) mediated by a physical agent (wind or water), or an animal, such as an insect or bird. Cross-pollination increases genetic diversity in a species.

Diversity of Angiosperms

Angiosperms are classified in a single phylum: the **Anthophyta**. Modern angiosperms appear to be a monophyletic group, which means that they originate from a single ancestor. Flowering plants are divided into two major groups, according to the structure of the cotyledons, pollen grains, and other structures. **Monocots** include grasses and lilies, and **eudicots** form a polyphyletic group. **Basal angiosperms** are a group of plants that are believed to have branched off before the separation into monocots and eudicots because they exhibit traits from both groups. They are categorized separately in many classification schemes. The *Magnoliidae* (magnolia trees, laurels, and water lilies) and the *Piperaceae* (peppers) belong to the basal angiosperm group.

Basal Angiosperms

The Magnoliidae are represented by the magnolias: tall trees bearing large, fragrant flowers that have many parts and are considered archaic (**Figure 21.41d**). Laurel trees produce fragrant leaves and small, inconspicuous flowers. The *Laurales* grow mostly in warmer climates and are small trees and shrubs. Familiar plants in this group include the bay laurel, cinnamon, spice bush (**Figure 21.41a**), and avocado tree. The *Nymphaeales* are comprised of the water lilies, lotus (**Figure 21.41c**), and similar plants; all species thrive in freshwater biomes, and have leaves that float on the water surface or grow underwater. Water lilies are particularly prized by gardeners, and have graced ponds and pools for thousands of years. The *Piperales* are a group of herbs, shrubs, and small trees that grow in the tropical climates. They have small flowers without petals that are tightly arranged in long spikes. Many species are the source of prized fragrance or spices, for example the berries of *Piper nigrum* (**Figure 21.41b**) are the familiar black peppercorns that are used to flavor many dishes.

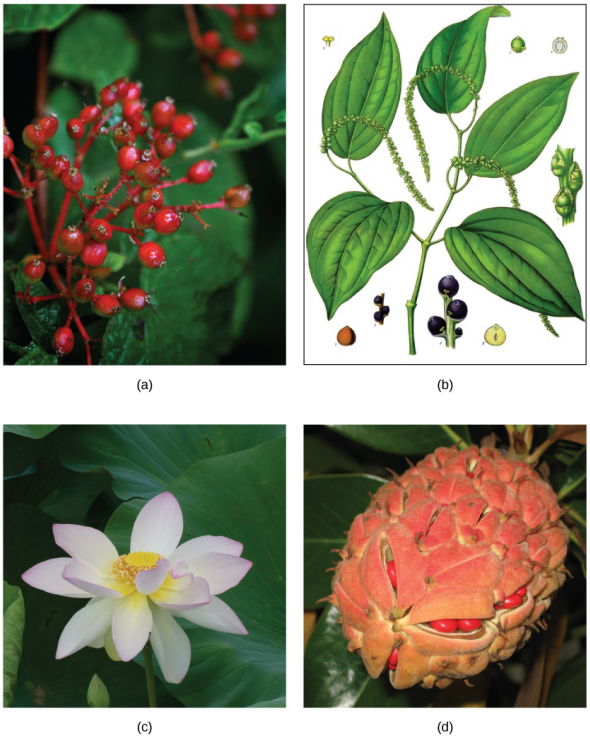


Figure 21.41 The (a) common spicebush belongs to the *Laurales*, the same family as cinnamon and bay laurel. The fruit of (b) the *Piper nigrum* plant is black pepper, the main product that was traded along spice routes. Notice the small, unobtrusive, clustered flowers. (c) Lotus flowers, *Nelumbo nucifera*, have been cultivated since ancient times for their ornamental value; the root of the lotus flower is eaten as a vegetable. The red seeds of (d) a magnolia tree, characteristic of the final stage, are just starting to appear. (credit a: modification of work by Cory Zanker; credit b: modification of work by Franz Eugen Köhler; credit c: modification of work by "berduchwal"/Flickr; credit d: modification of work by "Coastside2"/Wikimedia Commons).

Monocots

Plants in the monocot group are primarily identified as such by the presence of a single cotyledon in the seedling. Other anatomical features shared by monocots include veins that run parallel to the length of the leaves, and flower parts that are arranged in a three- or six-fold symmetry. In monocots, the vascular tissue is scattered in the stem. True woody tissue is rarely found in monocots. In palm trees, vascular and parenchyma tissues produced by the primary and secondary thickening meristems form the trunk. The pollen from the first angiosperms was monosulcate, containing a single furrow or pore through the outer layer. This feature is still seen in the modern monocots. Vascular tissue of the stem is not arranged in any particular pattern. The root system is mostly adventitious and unusually positioned, with no major tap root. The monocots include familiar plants such as the true lilies (which are at the origin of their alternate name of Liliopsida), orchids, grasses, and palms. Many important crops are monocots, such as rice and other cereals, corn, sugar cane, and tropical fruits like bananas and pineapples (**Figure 21.42**).

Figure 21.42 The world's major crops are flowering plants. (a) Rice, (b) wheat, and (c) bananas are monocots, while (d) cabbage, (e) beans, and (f) peaches are eudicots. (credit a: modification of work by David Nance, USDA ARS; credit b, c: modification of work by Rosendahl; credit d: modification of work by Bill Tarpenning, USDA; credit e: modification of work by Scott Bauer, USDA ARS; credit f: modification of work by Keith Weller, USDA)

Eudicots

Eudicots, or true dicots, are characterized by the presence of two cotyledons in the developing shoot. Veins form a network in leaves, and flower parts come in four, five, or many whorls. Vascular tissue forms a ring in the stem. Eudicots can be **herbaceous**, or produce woody tissues. Most eudicots produce pollen that is trisulcate or triporate, with three furrows or pores. The root system is usually anchored by one main root. Eudicots comprise two-thirds of all flowering plants. The major differences between monocots and eudicots are summarized in Table 21.1. Many species exhibit characteristics that belong to either group; as such, the classification of a plant as a monocot or a eudicot is not always clearly evident.

Comparison of Structural Characteristics of Monocots and Eudicots

Characteristic	Monocot	Eudicot
Cotyledon	One	Two
Veins in Leaves	Parallel	Network (branched)
Stem Vascular Tissue	Scattered	Arranged in ring pattern
Roots	Network of adventitious roots	Tap root with many lateral roots
Root Vascular Tissues	Xylem in ring pattern	Xylem in X or star pattern
Pollen	Single pore or furrow	Three pores or furrows
Flower Parts	Multiples of three	Multiples of four or five and whorls

Table 21.1

21.8 | Asexual Reproduction

Introduction

> “ From the sexual, or amatorial, generation of plants new varieties, or improvements, are frequently obtained; as many of the young plants from seeds are dissimilar to the parent, and some of them superior to the parent in the qualities we wish to possess... Sexual reproduction is the chef d'oeuvre, the master-piece of nature.”

Erasmus Darwin, English physician and grandfather of Charles Darwin, in *Phytologia*, (1800)

Sexual reproduction is certainly one way to produce the next generation, but it is not the only way. And in some situations, it might not be the best way. Many plants are able to propagate themselves using asexual reproduction by simply growing a new individual by mitosis. This method does not require the investment required to produce a flower, attract pollinators, or find a means of seed dispersal. Asexual reproduction produces plants that are genetically identical to the parent plant because no mixing of male and female gametes takes place. Traditionally, these plants survive well under stable environmental conditions when compared with plants produced from sexual reproduction because they carry genes identical to those of their parents.

Many different structures facilitate asexual reproduction (**Figure 21.43**). Garlic and gladiolus have an underground stem called a corm. Bulbs, such as a scaly bulb in lilies and a tunicate bulb in daffodils, are other common examples. A potato is a stem tuber, while parsnip propagates from a taproot. Ginger and iris produce rhizomes, while ivy uses an adventitious root (a root arising from a plant part other than the main or primary root), and the strawberry plant has a stolon, which is also called a runner. In each case, the tissue was grown via mitosis and becomes a new individual when separated from the 'parent' plant.

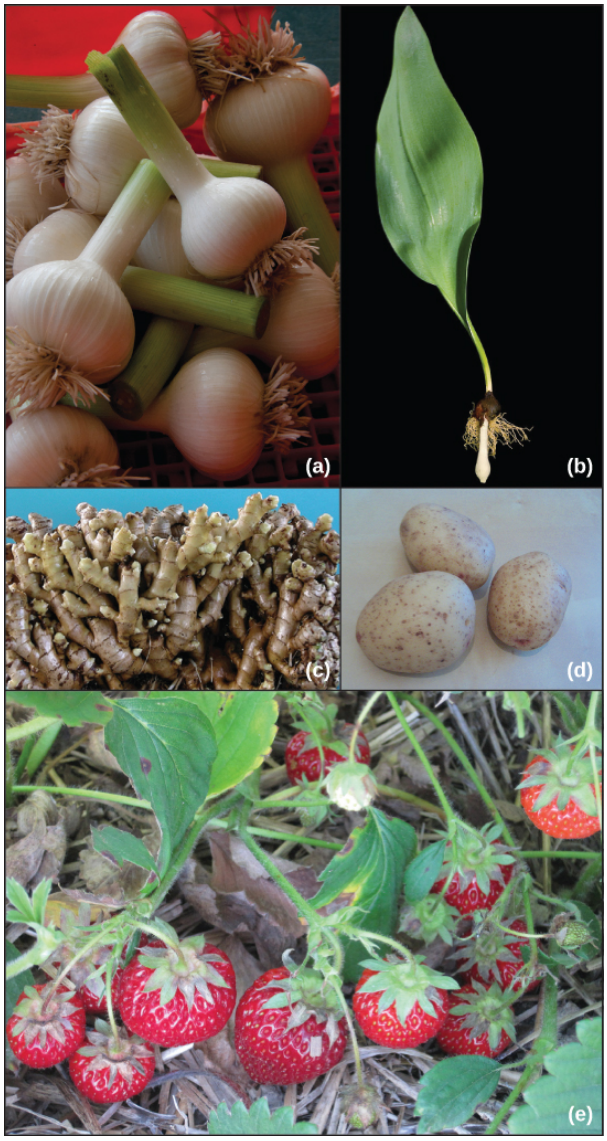


Figure 21.43 Different types of stems allow for asexual reproduction. (a) The corm of a garlic plant looks similar to (b) a tulip bulb, but the corm is solid tissue, while the bulb consists of layers of modified leaves that surround an underground stem. Both corms and bulbs can self-propagate, giving rise to new plants. (c) Ginger forms masses of stems called rhizomes that can give rise to multiple plants. (d) Potato plants form fleshy stem tubers. Each eye in the stem tuber can give rise to a new plant. (e) Strawberry plants form stolons: stems that grow at the soil surface or just below ground and can give rise to new plants. (credit a: modification of work by Dwight Sipler; credit c: modification of work by Albert Cahalan, USDA ARS; credit d: modification of work by Richard North; credit e: modification of work by Julie Magro)

Some plants can produce seeds without fertilization. Either the ovule or part of the ovary, which is diploid in nature, gives rise to a new seed. This method of reproduction is known as apomixis.

One advantage of asexual reproduction is that the resulting plant will reach maturity faster. Since the new plant is arising from an adult plant or plant parts, it will also be sturdier than a seedling. Asexual reproduction can take place by natural or artificial (assisted by humans) means.

Natural Methods of Asexual Reproduction

Natural methods of asexual reproduction include strategies listed above that plants have developed to self-propagate. Many plants—like ginger, onion, gladioli, and dahlia—continue to grow from buds that are present on the surface of the stem. In some plants, such as the sweet potato, adventitious roots or runners can give rise to new plants **Figure 21.44**. In *Bryophyllum* and kalanchoe, the leaves have small buds on their margins. When these are detached from the plant, they grow into independent plants; or, they may start growing into independent plants if the leaf touches the soil. Some plants can be propagated through cuttings alone.

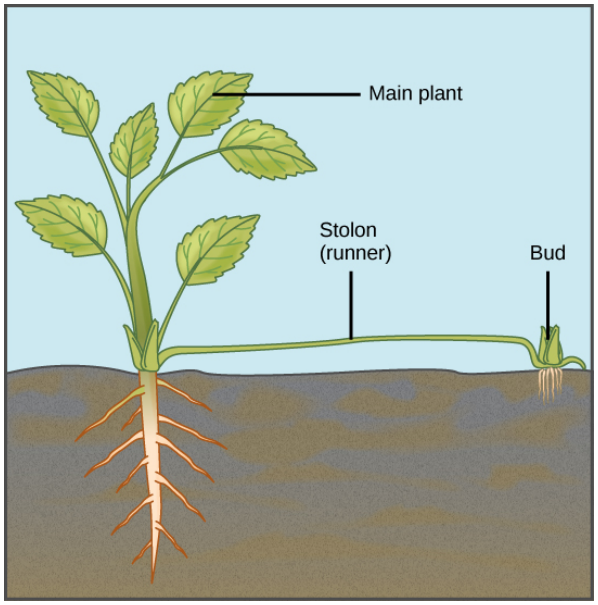


Figure 21.44 A stolon, or runner, is a stem that runs along the ground. At the nodes, it forms adventitious roots and buds that grow into a new plant.

Artificial Methods of Asexual Reproduction

These methods are frequently employed to give rise to new, and sometimes novel, plants. They include grafting, cutting, layering, and micropropagation.

Grafting

Grafting has long been used to produce novel varieties of roses, citrus species, and other plants. In **grafting**, two plant species are used; part of the stem of the desirable plant is grafted onto a rooted plant called the stock. The part that is grafted or attached is called the scion. Both are cut at an oblique angle (any angle other than a right angle), placed in close contact with each other, and are then held together **Figure 21.45**. Matching up these two surfaces as closely as possible is extremely important because these will be holding the plant together. The vascular systems of the two plants grow and fuse, forming a graft. After a period of time, the scion starts producing shoots, and eventually starts bearing flowers and fruits. Grafting is widely used in viticulture (grape growing) and the citrus industry. Scions capable of producing a particular fruit variety are grafted onto root stock with specific resistance to disease.

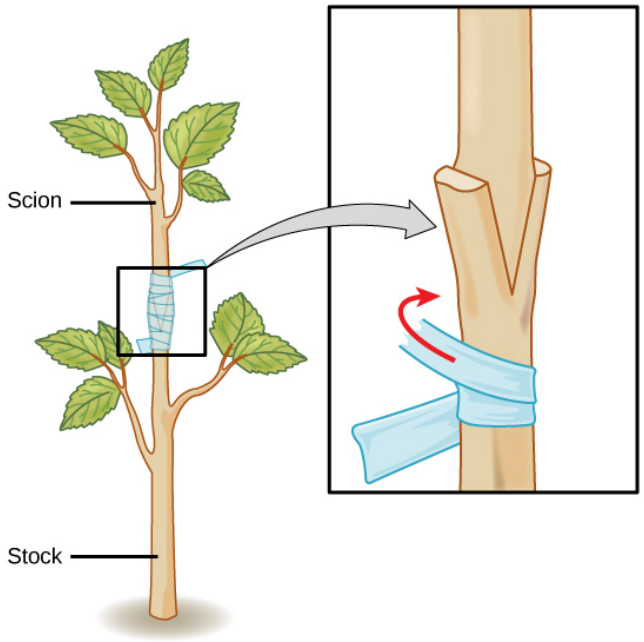


Figure 21.45 Grafting is an artificial method of asexual reproduction used to produce plants combining favorable stem characteristics with favorable root characteristics. The stem of the plant to be grafted is known as the scion, and the root is called the stock.

Cutting

Plants such as coleus and money plant are propagated through stem cuttings, where a portion of the stem containing nodes and internodes is placed in moist soil and allowed to root. In some species, stems can start producing a root even when placed only in water. For example, leaves of the African violet will root if kept in water undisturbed for several weeks.

Layering

Layering is a method in which a stem attached to the plant is bent and covered with soil. Young stems that can be bent easily without any injury are preferred. Jasmine and bougainvillea (paper flower) can be propagated this way **Figure 21.46**. In some plants, a modified form of layering known as air layering is employed. A portion of the bark or outermost covering of the stem is removed and covered with moss, which is then taped. Some gardeners also apply rooting hormone. After some time, roots will appear, and this portion of the plant can be removed and transplanted into a separate pot.

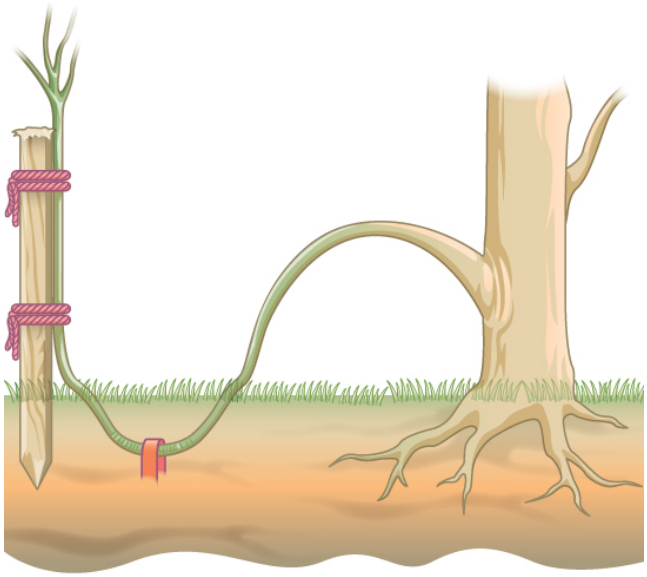


Figure 21.46 In layering, a part of the stem is buried so that it forms a new plant. (credit: modification of work by Pearson Scott Foresman, donated to the Wikimedia Foundation)

Micropropagation

Micropropagation (also called plant tissue culture) is a method of propagating a large number of plants from a single plant in a short time under laboratory conditions **Figure 21.47**. This method allows propagation of rare, endangered species that may be difficult to grow under natural conditions, are economically important, or are in demand as disease-free plants.

Figure 21.47 Micropropagation is used to propagate plants in sterile conditions. (credit: Nikhilesh Sanyal)

To start plant tissue culture, a part of the plant such as a stem, leaf, embryo, anther, or seed can be used. The plant material is thoroughly sterilized using a combination of chemical treatments standardized for that species. Under sterile conditions, the plant material is placed on a plant tissue culture medium that contains all the minerals, vitamins, and hormones required by the plant. The plant part often gives rise to an undifferentiated mass known as callus, from which individual plantlets begin to grow after a period of time. These can be separated and are first grown under greenhouse conditions before they are moved to field conditions.

Plant Life Spans

The length of time from the beginning of development to the death of a plant is called its life span. The life cycle, on the other hand, is the sequence of stages a plant goes through from seed germination to seed production of the mature plant. Some plants, such as annuals, only need a few weeks to grow, produce seeds and die. Other plants, such as the bristlecone pine, live for thousands of years. Some bristlecone pines have a documented age of 4,500 years **Figure 21.48**. Even as some parts of a plant, such as regions containing meristematic tissue—the area of active plant growth consisting of undifferentiated cells capable of cell division—continue to grow, some parts undergo programmed cell death (apoptosis). The cork found on stems, and the water-conducting tissue of the xylem, for example, are composed of dead cells.

Figure 21.48 The bristlecone pine, shown here in the Ancient Bristlecone Pine Forest in the White Mountains of eastern California, has been known to live for 4,500 years. (credit: Rick Goldwaser)

Plant species that complete their lifecycle in one season are known as annuals, an example of which is *Arabidopsis*, or mouse-ear cress. Biennials such as carrots complete their lifecycle in two seasons. In a biennial's first season, the plant has a vegetative phase, whereas in the next season, it completes its reproductive phase. Commercial growers harvest the carrot roots after the first year of growth, and do not allow the plants to flower. Perennials, such as the magnolia, complete their lifecycle in two years or more.

In another classification based on flowering frequency, monocarpic plants flower only once in their lifetime; examples include bamboo and yucca. During the vegetative period of their life cycle (which may be as long as 120 years in some bamboo species), these plants may reproduce asexually and accumulate a great deal of food material that will be required during their once-in-a-lifetime flowering and setting of seed after fertilization. Soon after flowering, these plants die. Polycarpic plants form flowers many times during their lifetime. Fruit trees, such as apple and orange trees, are polycarpic; they flower every year. Other polycarpic species, such as perennials, flower several times during their life span, but not each year. By this means, the plant does not require all its nutrients to be channeled towards flowering each year.

As is the case with all living organisms, genetics and environmental conditions have a role to play in determining how long a plant will live. Susceptibility to disease, changing environmental conditions, drought, cold, and competition for nutrients are some of the factors that determine the survival of a plant. Plants continue to grow, despite the presence of dead tissue such as cork. Individual parts of plants, such as flowers and leaves, have different rates of survival. In many trees, the older leaves turn yellow and eventually fall from the tree. Leaf fall is triggered by factors such as a decrease in photosynthetic efficiency, due to shading by upper leaves, or oxidative damage incurred as a result of photosynthetic reactions. The components of the part to be shed are recycled by the plant for use in other processes, such as development of seed and storage. This process is known as nutrient recycling.

The aging of a plant and all the associated processes is known as senescence, which is marked by several complex biochemical changes. One of the characteristics of senescence is the breakdown of chloroplasts, which is characterized by the yellowing of leaves. The chloroplasts contain components of photosynthetic machinery such as membranes and proteins. Chloroplasts also contain DNA. The proteins, lipids, and nucleic acids are broken down by specific enzymes into smaller molecules and salvaged by the plant to support the growth of other plant tissues.

The complex pathways of nutrient recycling within a plant are not well understood. Hormones are known to play a role in senescence. Applications of cytokinins and ethylene delay or prevent senescence; in contrast, abscissic acid causes premature onset of senescence.

22 | PLANT REPRODUCTION; STRUCTURE AND FUNCTION OF PLANT TISSUES

22.1 | The Plant Body

Introduction

> “ A mile and a half from town, I came to a grove of tall cocoanut trees, with clean, branchless stems reaching straight up sixty or seventy feet and topped with a spray of green foliage sheltering clusters of cocoanuts—not more picturesque than a forest of colossal ragged parasols, with bunches of magnified grapes under them, would be. I once heard a grouty northern invalid say that a cocoanut tree might be poetical, possibly it was; but it looked like a feather-duster struck by lightning. I think that describes it better than a picture—and yet, without any question, there is something fascinating about a cocoanut tree—and graceful, too.”

Mark Twain, *Roughing it*, (1913)

Figure 22.1 Palm tree on a tropical beach Palm tree on a beach, Tulum, Mexico. (Photo by David A. Rintoul)

Twain's observations about the "cocoanut tree" focuses on some of the more obvious parts of a plant: the stem, the leaves, the seeds, etc. Some of the other parts of a plant are less obvious, but not less important to the health and well-being of the organism. Like animals, plants have organelles, cells, tissues and organs.

Plant Organ Systems

In plants, just as in animals, similar cells working together form a tissue. When different types of tissues work together to perform a unique function, they form an organ; organs working together form organ systems. Vascular plants have two distinct organ systems: a shoot system, and a root system. The **shoot system** consists of two portions: the vegetative (non-reproductive) parts of the plant, such as the leaves and the stems, and the reproductive parts of the plant, which include flowers and fruits. The shoot system generally grows above ground, where it absorbs the light needed for photosynthesis. The **root system**, which supports the plants and absorbs water and minerals, is usually underground. **Figure 22.2** shows the organ systems of a typical plant.

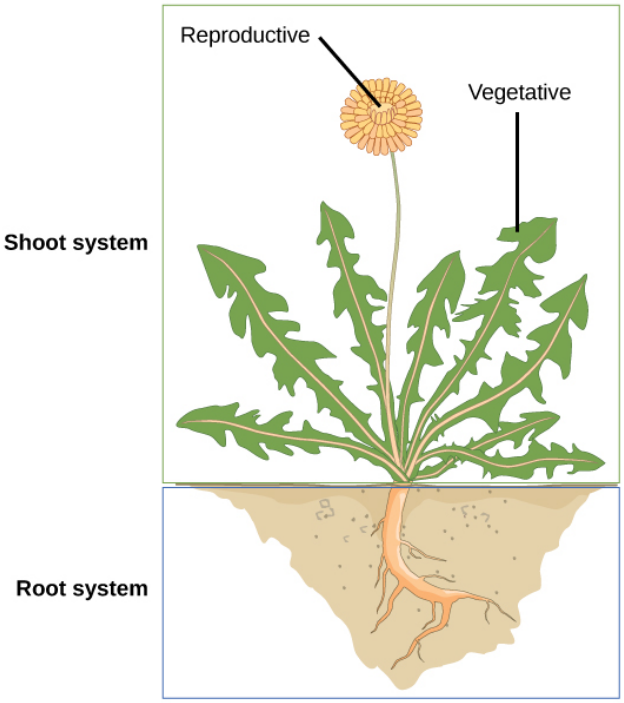


Figure 22.2 The shoot system of a plant consists of leaves, stems, flowers, and fruits. The root system anchors the plant while absorbing water and minerals from the soil.

Plant Tissues

Plants are multicellular eukaryotes with tissue systems made of various cell types that carry out specific functions. Plant tissue systems fall into one of two general types: meristematic tissue, and permanent (or non-meristematic) tissue. Cells of the meristematic tissue are found in **meristems**, which are plant regions of continuous cell division and growth. **Meristematic tissue** cells are either undifferentiated or incompletely differentiated, and they continue to divide and contribute to the growth of the plant. In contrast, permanent tissue consists of plant cells that are no longer actively dividing.

Meristematic tissues consist of three types, based on their location in the plant. **Apical meristems** contain meristematic tissue located at the tips of stems and roots, which enable a plant to extend in length. **Lateral meristems** facilitate growth in thickness or girth in a maturing plant. Intercalary meristems occur only in monocots, at the bases of leaf blades and at nodes (the areas where leaves attach to a stem). This tissue enables the monocot leaf blade to increase in length from the leaf base; for example, it allows lawn grass leaves to elongate even after repeated mowing.

Meristems produce cells that quickly differentiate, or specialize, and become permanent tissue. Such cells take on specific roles and lose their ability to divide further. They differentiate into three main types: dermal, vascular, and ground tissue. **Dermal tissue** covers and protects the plant, and **vascular tissue** transports water, minerals, and sugars to different parts of the plant. **Ground tissue** serves as a site for photosynthesis, provides a supporting matrix for the vascular tissue, and helps to store water and sugars.

Secondary tissues are either simple (composed of similar cell types) or complex (composed of different cell types). Dermal tissue, for example, is a simple tissue that covers the outer surface of the plant and controls gas exchange. Vascular tissue is an example of a complex tissue, and is made of two specialized conducting tissues: xylem and phloem. Xylem tissue transports water and nutrients from the roots to different parts of the plant, and includes three different cell types: vessel elements and tracheids (both of which conduct water), and xylem parenchyma. Phloem tissue, which transports organic compounds from the site of photosynthesis to other parts of the plant, consists of four different cell types: sieve cells (which conduct photosynthates), companion cells, phloem parenchyma, and phloem fibers. Unlike xylem conducting cells, phloem conducting cells are alive at maturity. The xylem and phloem always lie adjacent to each other (**Figure 22.3**). In stems, the xylem and the phloem form a structure called a **vascular bundle**; in roots, this is termed the vascular cylinder.

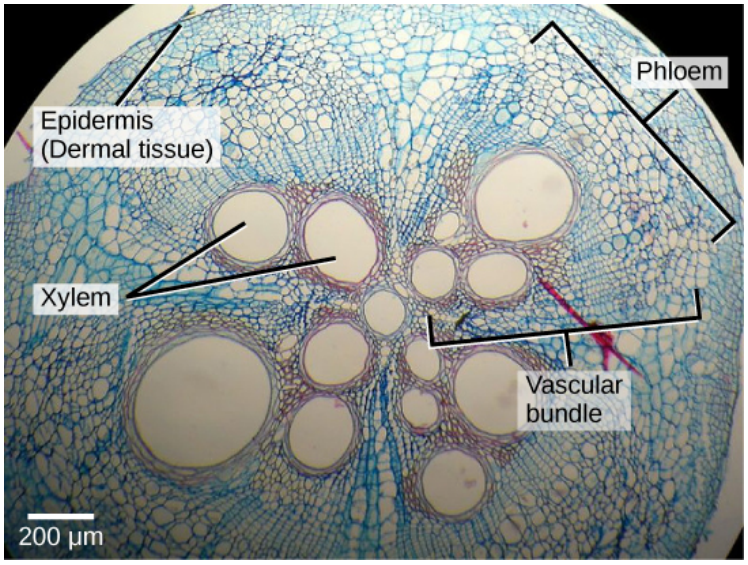


Figure 22.3 This light micrograph shows a cross section of a squash (*Curcurbita maxima*) stem. Each teardrop-shaped vascular bundle consists of large xylem vessels toward the inside and smaller phloem cells toward the outside. Xylem cells, which transport water and nutrients from the roots to the rest of the plant, are dead at functional maturity. Phloem cells, which transport sugars and other organic compounds from photosynthetic tissue to the rest of the plant, are living. The vascular bundles are encased in ground tissue and surrounded by dermal tissue. (credit: modification of work by "(biophotos)"/Flickr; scale-bar data from Matt Russell)

Stems

Stems are a part of the shoot system of a plant. They may range in length from a few millimeters to hundreds of meters, and also vary in diameter, depending on the plant type. Stems are usually above ground, although the stems of some plants, such as the potato, also grow underground. Stems may be herbaceous (soft) or woody in nature. Their main function is to provide support to the plant, holding leaves, flowers and buds; in some cases, stems also store food for the plant. A stem may be unbranched, like that of a palm tree, or it may be highly branched, like that of an oak or pine tree. The stem of the plant connects the roots to the leaves, helping to transport absorbed water and minerals to different parts of the plant. It also helps to transport the products of photosynthesis, namely sugars, from the leaves to the rest of the plant.

Plant stems, whether above or below ground, are characterized by the presence of nodes and internodes (**Figure 22.4**). Nodes are points of attachment for leaves, aerial roots, and flowers. The stem region between two nodes is called an internode. The stalk that extends from the stem to the base of the leaf is the petiole. An axillary bud is usually found in the axil—the area between the base of a leaf and the stem—where it can give rise to a branch or a flower. The apex (tip) of the shoot contains the apical meristem within the apical bud.

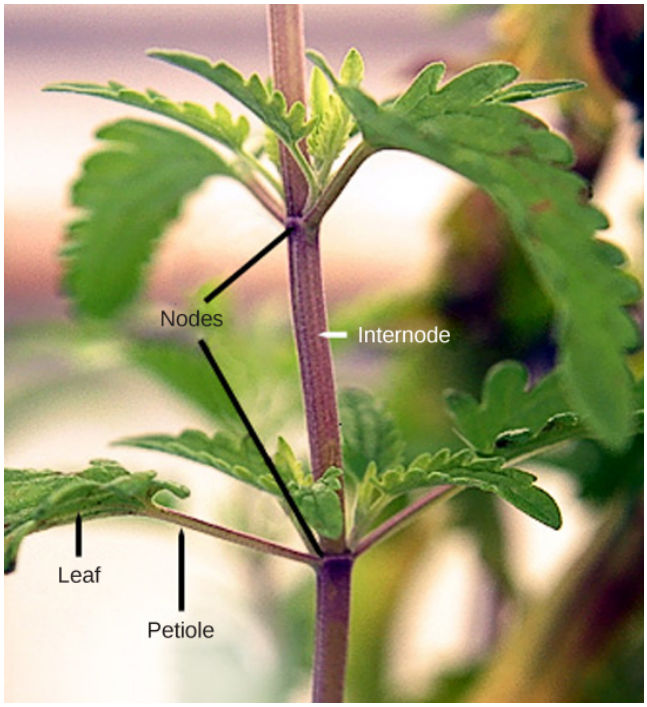


Figure 22.4 Leaves are attached to the plant stem at areas called nodes. An internode is the stem region between two nodes. The petiole is the stalk connecting the leaf to the stem. The leaves just above the nodes arose from axillary buds.

Stem Anatomy

The stem and other plant organs arise from the ground tissue, and are primarily made up of simple tissues formed from three types of cells: parenchyma, collenchyma, and sclerenchyma cells.

Parenchyma cells are the most common plant cells (**Figure 22.5**). They are found in the stem, the root, the inside of the leaf, and the pulp of the fruit. Parenchyma cells are responsible for metabolic functions, such as photosynthesis, and they help repair and heal wounds. Some parenchyma cells also store starch.

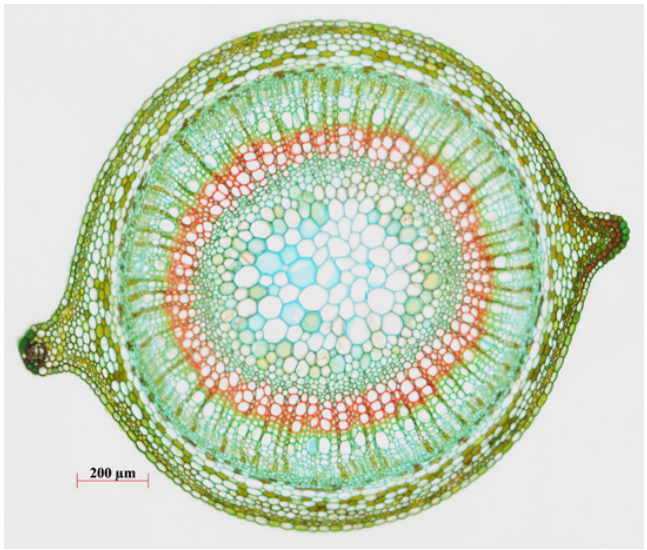


Figure 22.5 The stem of common St John's Wort (*Hypericum perforatum*) is shown in cross section in this light micrograph. The central pith (greenish-blue, in the center) and peripheral cortex (narrow zone 3–5 cells thick just inside the epidermis) are composed of parenchyma cells. Vascular tissue composed of xylem (red) and phloem tissue (green, between the xylem and cortex) surrounds the pith. (credit: Rolf-Dieter Mueller)

Collenchyma cells are elongated cells with unevenly thickened walls (**Figure 22.6**). They provide structural support, mainly to the stem and leaves. These cells are alive at maturity and are usually found below the epidermis. The “strings” of a celery

stalk are an example of collenchyma cells.

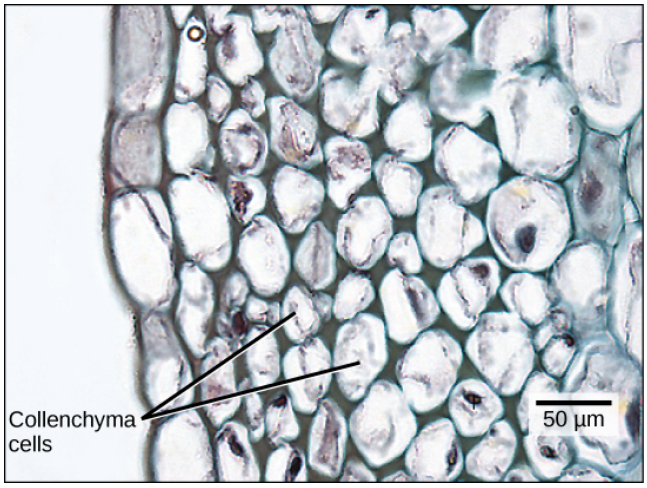


Figure 22.6 Collenchyma cell walls are uneven in thickness, as seen in this light micrograph. They provide support to plant structures. (credit: modification of work by Carl Szczerski; scale-bar data from Matt Russell)

Sclerenchyma cells also provide support to the plant, but unlike collenchyma cells, many of them are dead at maturity. There are two types of sclerenchyma cells: fibers and sclereids. Both types have secondary cell walls that are thickened with deposits of lignin, an organic compound that is a key component of wood. Fibers are long, slender cells; sclereids are smaller-sized. Sclereids give pears their gritty texture. Humans use sclerenchyma fibers to make linen and rope (**Figure 22.7**).

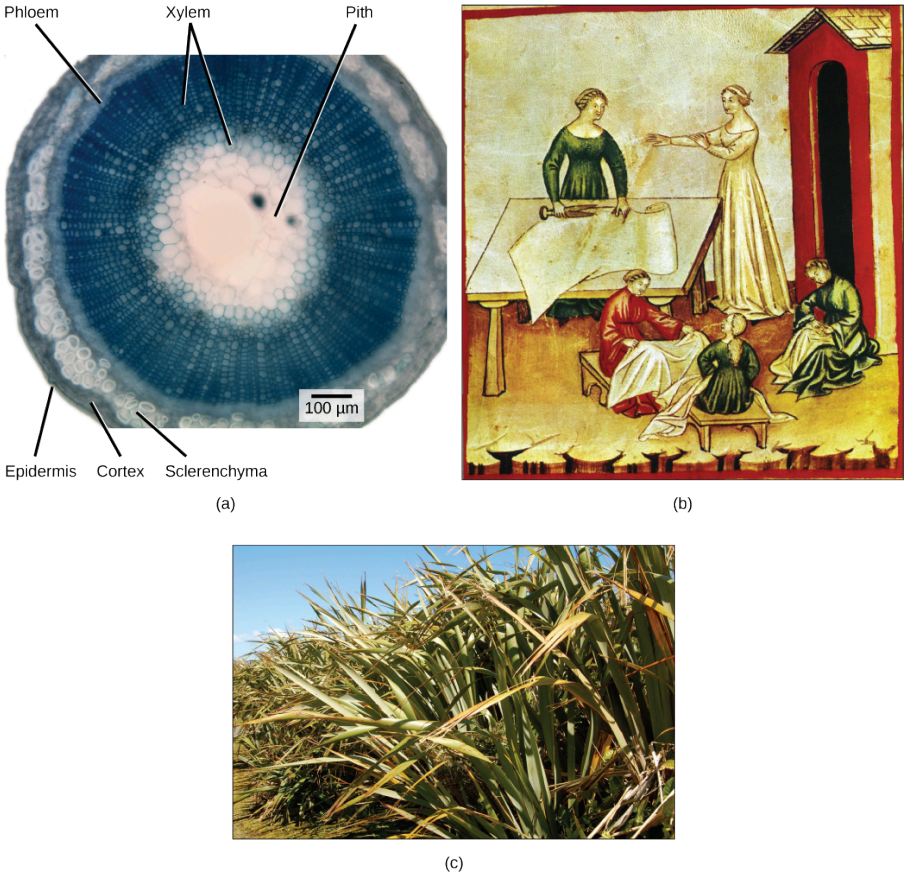


Figure 22.7 The central pith and outer cortex of the (a) flax stem are made up of parenchyma cells. Inside the cortex is a layer of sclerenchyma cells, which make up the fibers in flax rope and clothing. Humans have grown and harvested flax for thousands of years. In (b) this drawing, fourteenth-century women prepare linen. The (c) flax plant is grown and harvested for its fibers, which are used to weave linen, and for its seeds, which are the source of linseed oil. (credit a: modification of work by Emmanuel Boutet based on original work by Ryan R. MacKenzie; credit c: modification of work by Brian Dearth; scale-bar data from Matt Russell)

Like the rest of the plant, the stem has three tissue systems: dermal, vascular, and ground tissue. Each is distinguished by characteristic cell types that perform specific tasks necessary for the plant's growth and survival.

Dermal Tissue

The dermal tissue of the stem consists primarily of epidermis, a single layer of cells covering and protecting the underlying tissue. Woody plants have a tough, waterproof outer layer of cork cells commonly known as bark, which further protects the plant from damage. Epidermal cells are the most numerous and least differentiated of the cells in the epidermis. The epidermis of a leaf also contains openings known as stomata, through which the exchange of gases takes place (**Figure 22.8**). Two cells, known as guard cells, surround each leaf stoma, controlling its opening and closing and thus regulating the uptake of carbon dioxide and the release of oxygen and water vapor. Trichomes are hair-like structures on the epidermal surface. They help to reduce **transpiration** (the loss of water by aboveground plant parts), increase solar reflectance, and store compounds that defend the leaves against predation by herbivores.

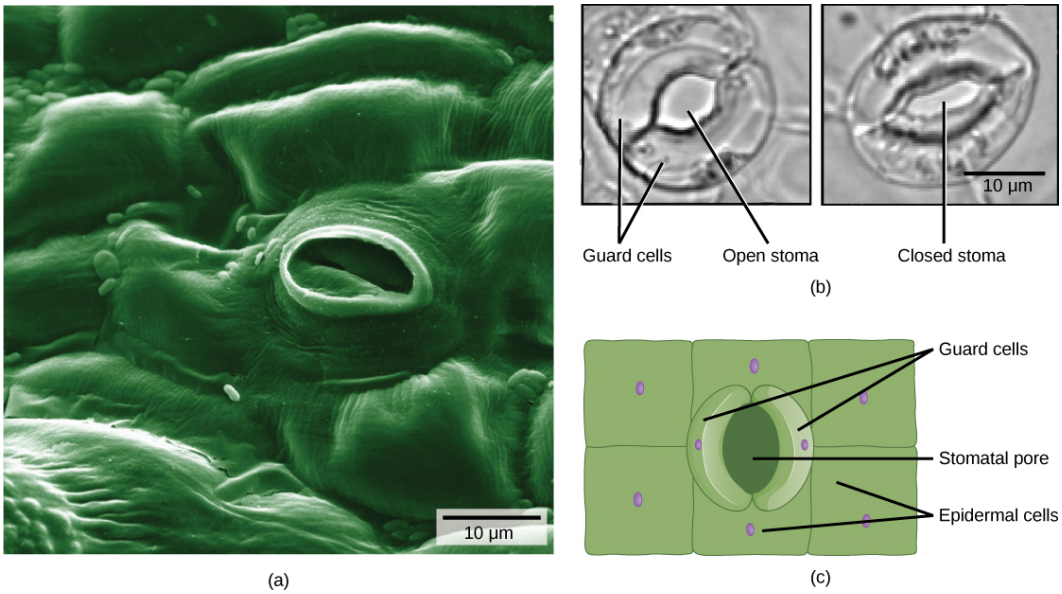


Figure 22.8 Openings called stomata (singular: stoma) allow a plant to take up carbon dioxide and release oxygen and water vapor. The (a) colorized scanning-electron micrograph shows a closed stoma of a eudicot. Each stoma is flanked by two guard cells that regulate its (b) opening and closing. The (c) guard cells sit within the layer of epidermal cells (credit a: modification of work by Louisa Howard, Rippel Electron Microscope Facility, Dartmouth College; credit b: modification of work by June Kwak, University of Maryland; scale-bar data from Matt Russell)

Vascular Tissue

The xylem and phloem that make up the vascular tissue of the stem are arranged in distinct strands called vascular bundles, which run up and down the length of the stem. When the stem is viewed in cross section, the vascular bundles of eudicot stems are arranged in a ring. In plants with stems that live for more than one year, the individual bundles grow together and produce the characteristic growth rings. In monocot stems, the vascular bundles are randomly scattered throughout the ground tissue (**Figure 22.9**).

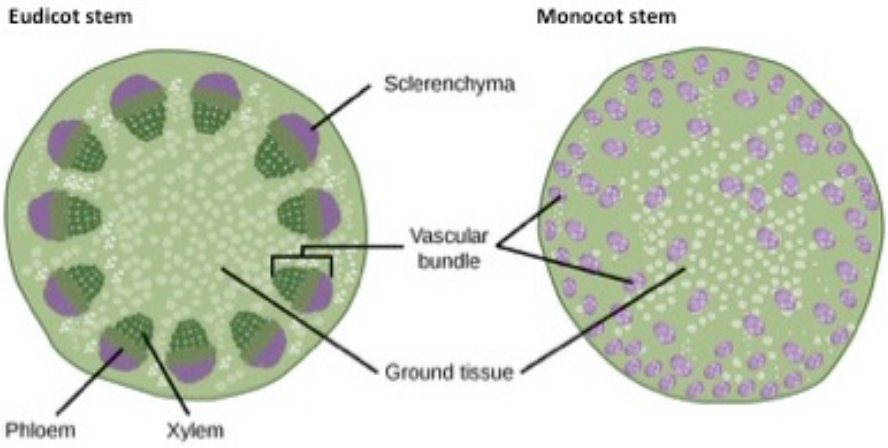


Figure 22.9 In (a) eudicot stems, vascular bundles are arranged around the periphery of the ground tissue. The xylem tissue is located toward the interior of the vascular bundle, and phloem is located toward the exterior. Sclerenchyma fibers cap the vascular bundles. In (b) monocot stems, vascular bundles composed of xylem and phloem tissues are scattered throughout the ground tissue.

Xylem tissue has three types of cells: xylem parenchyma, tracheids, and vessel elements. The latter two types conduct water and are dead at maturity. Tracheids are xylem cells with thick secondary cell walls that are lignified. Water moves from one tracheid to another through regions on the side walls known as pits, where secondary walls are absent. Vessel elements are xylem cells with thinner walls; they are shorter than tracheids. Each vessel element is connected to the next by means of a

perforation plate at the end walls of the element. Water moves through the perforation plates to travel up the plant.

Phloem tissue is composed of sieve-tube cells, companion cells, phloem parenchyma, and phloem fibers. A series of sieve-tube cells (also called sieve-tube elements) are arranged end to end to make up a long sieve tube, which transports organic substances such as sugars and amino acids. The sugars flow from one sieve-tube cell to the next through perforated sieve plates, which are found at the end junctions between two cells. Although still alive at maturity, the nucleus and other cell components of the sieve-tube cells have disintegrated. Companion cells are found alongside the sieve-tube cells, providing them with metabolic support. The companion cells contain more ribosomes and mitochondria than the sieve-tube cells, which lack some cellular organelles.

Ground Tissue

Ground tissue is mostly made up of parenchyma cells, but may also contain collenchyma and sclerenchyma cells that help support the stem. The ground tissue towards the interior of the vascular tissue in a stem or root is known as **pith**, while the layer of tissue between the vascular tissue and the epidermis is known as the **cortex**.

Growth in Stems

Growth in plants occurs as the stems and roots lengthen. Some plants, especially those that are woody, also increase in thickness during their life span. The increase in length of the shoot and the root is referred to as **primary growth**, and is the result of cell division in the shoot apical meristem. **Secondary growth** is characterized by an increase in thickness or girth of the plant, and is caused by cell division in the lateral meristem. Figure 22.10 shows the areas of primary and secondary growth in a plant. Herbaceous plants mostly undergo primary growth, with hardly any secondary growth or increase in thickness. Secondary growth or "wood" is noticeable in woody plants; it occurs in some eudicots, but occurs very rarely in monocots.

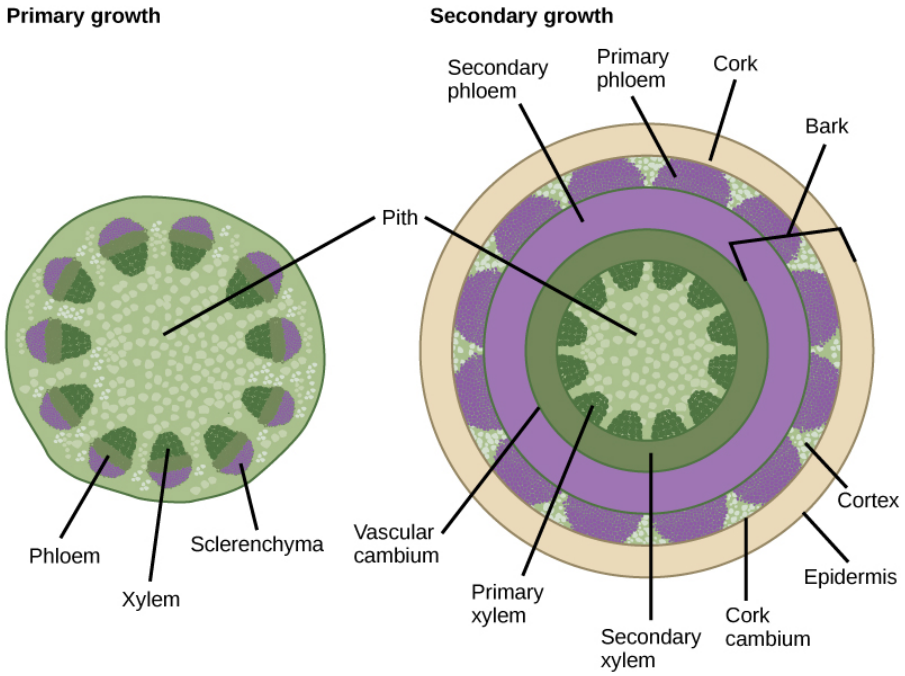


Figure 22.10 In woody plants, primary growth is followed by secondary growth, which allows the plant stem to increase in thickness or girth. Secondary vascular tissue is added as the plant grows, as well as a cork layer. The bark of a tree extends from the vascular cambium to the epidermis.

Some plant parts, such as stems and roots, continue to grow throughout a plant's life: a phenomenon called indeterminate growth. Other plant parts, such as leaves and flowers, exhibit determinate growth, which ceases when a plant part reaches a particular size.

Annual Rings

The activity of the vascular cambium gives rise to annual growth rings. During the spring growing season, cells of the

secondary xylem have a large internal diameter and their primary cell walls are not extensively thickened. This is known as early wood, or spring wood. During the fall season, the secondary xylem develops thickened cell walls, forming late wood, or autumn wood, which is denser than early wood. This alternation of early and late wood is due largely to a seasonal decrease in the number of vessel elements and a seasonal increase in the number of tracheids. It results in the formation of an annual ring, which can be seen as a circular ring in the cross section of the stem (**Figure 22.11**). An examination of the number of annual rings and their nature (such as their size and cell wall thickness) can reveal the age of the tree and the prevailing climatic conditions during each season.

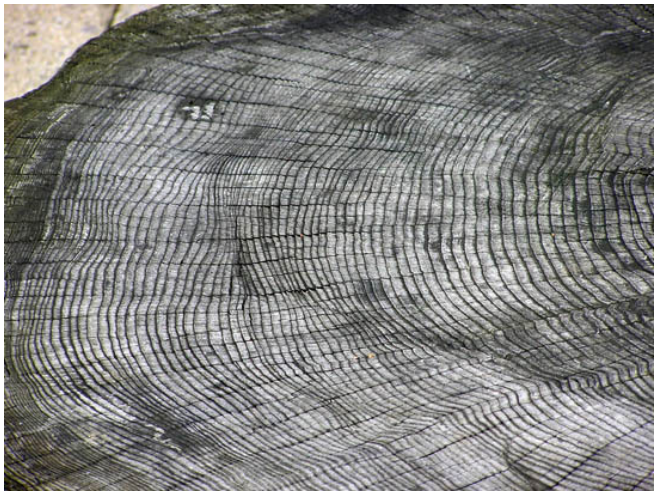


Figure 22.11 The rate of wood growth increases in summer and decreases in winter, producing a characteristic ring for each year of growth. Seasonal changes in weather patterns can also affect the growth rate—note how the rings vary in thickness. (credit: Adrian Pingstone)

Roots

The roots of seed plants have three major functions: anchoring the plant to the soil, absorbing water and minerals and transporting them upwards, and storing the products of photosynthesis. Some roots are modified to absorb moisture and exchange gases. Most roots are underground. Some plants, however, also have adventitious roots, which emerge above the ground from the shoot.

Types of Root Systems

Root systems are mainly of two types (**Figure 22.12**). Eudicots have a tap root system, while monocots have a fibrous root system. A **tap root system** has a main root that grows down vertically, and from which many smaller lateral roots arise. Dandelions are a good example; their tap roots usually break off when trying to pull these weeds, and they can regrow another shoot from the remaining root). A tap root system penetrates deep into the soil. In contrast, a **fibrous root system** is located closer to the soil surface, and forms a dense network of roots that also helps prevent soil erosion (lawn grasses are a good example, as are wheat, rice, and corn). Some plants have a combination of tap roots and fibrous roots. Plants that grow in dry areas often have deep root systems, whereas plants growing in areas with abundant water are likely to have shallower root systems.

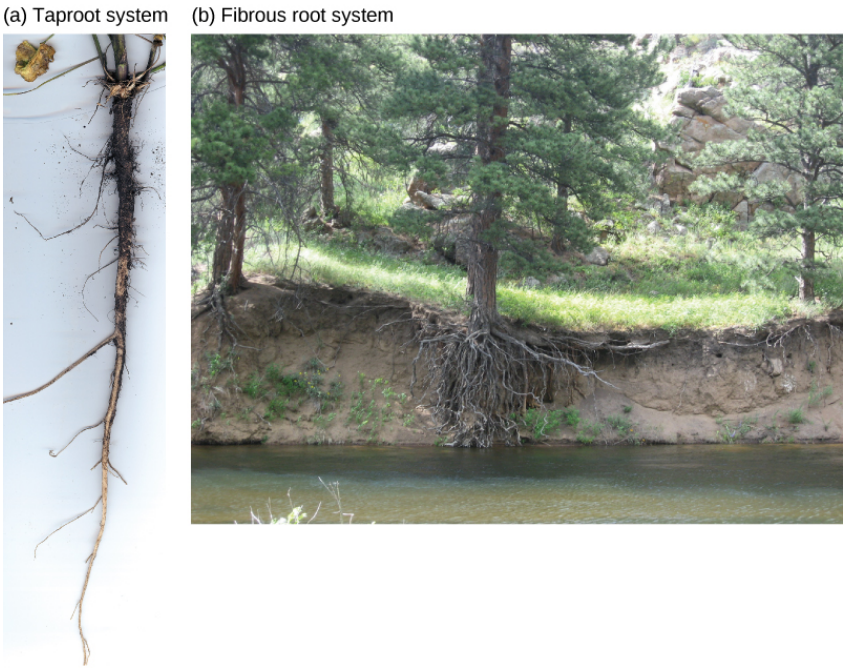


Figure 22.12 (a) Tap root systems have a main root that grows down, while (b) fibrous root systems consist of many small roots. (credit b: modification of work by “Austen Squarepants”/Flickr)

Leaves

Leaves are the main sites for photosynthesis: the process by which plants synthesize food. Most leaves are usually green, due to the presence of chlorophyll in the leaf cells. However, some leaves may have different colors, caused by other plant pigments that mask the green chlorophyll.

The thickness, shape, and size of leaves are adapted to the environment. Each variation helps a plant species maximize its chances of survival in a particular habitat. Usually, the leaves of plants growing in tropical rainforests have larger surface areas than those of plants growing in deserts or very cold conditions, which are likely to have a smaller surface area to minimize water loss.

Structure of a Typical Leaf

Each leaf typically has a leaf blade called the lamina, which is also the widest part of the leaf. Some leaves are attached to the plant stem by a petiole. Leaves that do not have a petiole and are directly attached to the plant stem are called sessile leaves. Small green appendages usually found at the base of the petiole are known as stipules. Most leaves have a midrib, which travels the length of the leaf and branches to each side to produce veins of vascular tissue. The edge of the leaf is called the margin. **Figure 22.13** shows the structure of a typical eudicot leaf.

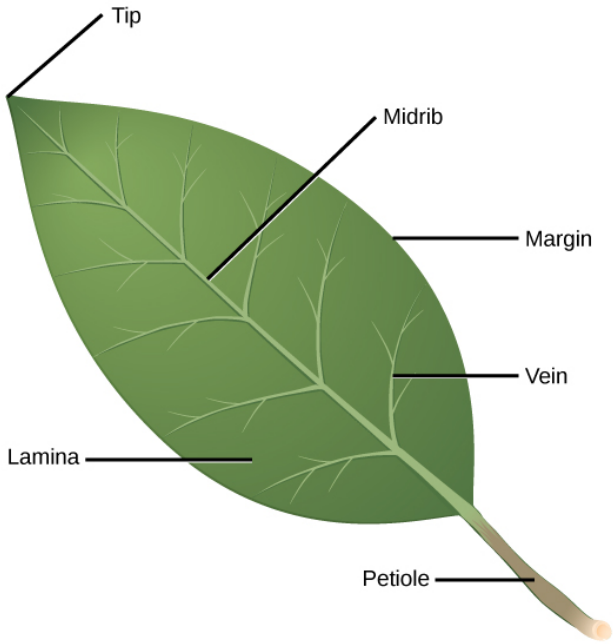


Figure 22.13 Deceptively simple in appearance, a leaf is a highly efficient structure.

Within each leaf, the vascular tissue forms veins. The arrangement of veins in a leaf is called the venation pattern. Monocots and eudicots differ in their patterns of venation (**Figure 22.14**). Monocots have parallel venation; the veins run in straight lines across the length of the leaf without converging at a point. In eudicots, however, the veins of the leaf have a net-like appearance, forming a pattern known as reticulate venation. One extant plant, the *Ginkgo biloba*, has dichotomous venation where the veins fork.

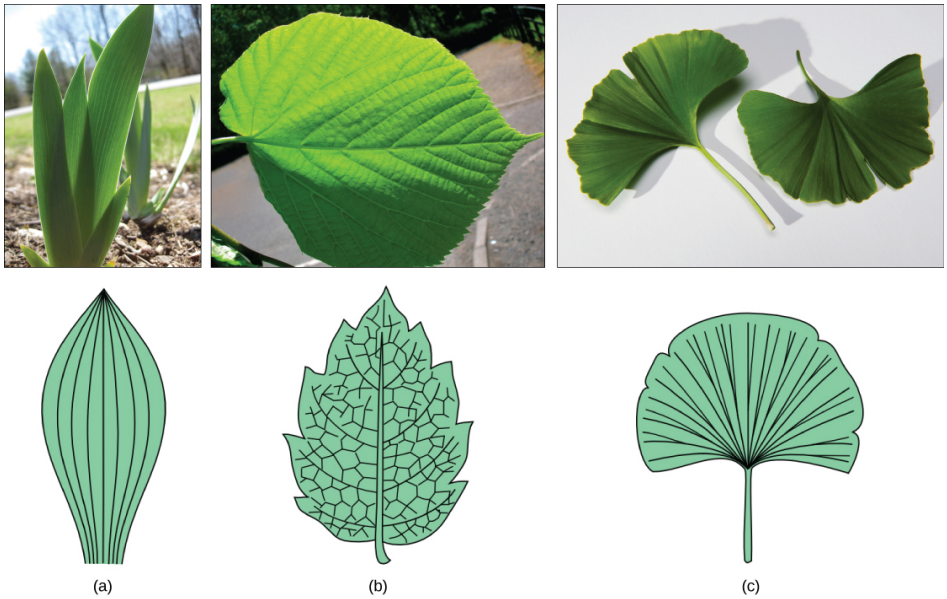


Figure 22.14 (a) Tulip (*Tulipa*), a monocot, has leaves with parallel venation. The netlike venation in this (b) linden (*Tilia cordata*) leaf distinguishes it as a eudicot. The (c) *Ginkgo biloba* tree has dichotomous venation. (credit a photo: modification of work by “Drewboy64”/Wikimedia Commons; credit b photo: modification of work by Roger Griffith; credit c photo: modification of work by "geishaboy500"/Flickr; credit abc illustrations: modification of work by Agnieszka Kwiecień)

22.2 | Plant Sensory Systems and Responses

Introduction

> "Plants are extraordinary. For instance ... if you pinch a leaf of a plant you set off electrical impulse. You can't touch a plant without setting off an electrical impulse ... There is no question that plants have all kinds of sensitivities. They do a lot of responding to an environment. They can do almost anything you can think of."

Barbara McClintock, Nobel prize-winning geneticist, in George Ritzer and Barry Smart, *Handbook of Social Theory* (2001)

Animals can respond to environmental factors by moving to a new location. Plants, however, are rooted in place and must respond to the surrounding environmental factors. Plants have sophisticated systems to detect and respond to light, gravity, temperature, and physical touch. Receptors sense environmental factors and relay the information to effector systems—often through intermediate chemical messengers—to bring about plant responses.

Plant Responses to Light

Plants have a number of sophisticated uses for light that go far beyond their ability to photosynthesize low-molecular-weight sugars using only carbon dioxide, light, and water. Photomorphogenesis is the growth and development of plants in response to light. It allows plants to optimize their use of light and space. Photoperiodism is the ability to use light to track time. Plants can tell the time of day and time of year by sensing and using various wavelengths of sunlight. **Phototropism** is a directional response that allows plants to grow towards, or even away from, light.

The sensing of light in the environment is important to plants; it can be crucial for competition and survival. The response of plants to light is mediated by different photoreceptors, which are comprised of a protein covalently bonded to a light-absorbing pigment called a chromophore. Together, the two are called a chromoprotein.

Phototropism

Phototropism—the directional bending of a plant toward or away from a light source—is a response to blue wavelengths of light. Positive phototropism is growth towards a light source (**m47403 (http://legacy.cnx.org/content/m47403/1.6/#figure_30_06_02)**), while negative phototropism is growth away from light.

In their 1880 treatise *The Power of Movements in Plants*, Charles Darwin and his son Francis first described phototropism as the bending of seedlings toward light. Darwin observed that light was perceived by the tip of the plant (the apical meristem), but that the response (bending) took place in a different part of the plant. They concluded that the signal had to travel from the apical meristem to the base of the plant.

Figure 22.15 Sunflowers in a field near Hutchinson, KS display a phototropic response by bending toward the light. (credit: David A. Rintoul)

In 1913, Peter Boysen-Jensen demonstrated that a chemical signal produced in the plant tip was responsible for the bending at the base. He cut off the tip of a seedling, covered the cut section with a layer of gelatin, and then replaced the tip. The seedling bent toward the light when illuminated. However, when impermeable mica flakes were inserted between the tip and

the cut base, the seedling did not bend. A refinement of the experiment showed that the signal traveled on the shaded side of the seedling. When the mica plate was inserted on the illuminated side, the plant did bend towards the light. Therefore, the chemical signal was a growth stimulant because the phototropic response involved faster cell elongation on the shaded side than on the illuminated side. We now know that the plant hormone auxin influences plant stem cell elongation, and accumulates on the shaded side of stems.

Plant Responses to Gravity

Whether or not they germinate in the light or in total darkness, shoots usually sprout up from the ground, and roots grow downward into the ground. A plant laid on its side in the dark will send shoots upward when given enough time. Gravitropism ensures that roots grow into the soil and that shoots grow toward sunlight. Growth of the shoot apical tip upward is called **negative gravitropism**, whereas growth of the roots downward is called **positive gravitropism**.

Amyloplasts (also known as statoliths) are specialized plastids that contain starch granules and settle downward in response to gravity. Amyloplasts are found in shoots and in specialized cells of the root cap. When a plant is tilted, the statoliths drop to the new bottom cell wall. A few hours later, the shoot or root will show growth in the new vertical direction.

The mechanism that mediates gravitropism is reasonably well understood. When amyloplasts settle to the bottom of the gravity-sensing cells in the root or shoot, they physically contact the endoplasmic reticulum (ER), causing the release of calcium ions from inside the ER. This calcium signaling in the cells causes polar transport of the plant hormone auxin to the bottom of the cell. In roots, a high concentration of auxin inhibits cell elongation. The effect slows growth on the lower side of the root, while cells develop normally on the upper side. Auxin has the opposite effect in shoots, where a higher concentration at the lower side of the shoot stimulates cell expansion, causing the shoot to grow up. After the shoot or root begin to grow vertically, the amyloplasts return to their normal position. Other hypotheses—involving the entire cell in the gravitropism effect—have been proposed to explain why some mutants that lack amyloplasts may still exhibit a weak gravitropic response.

Growth Responses

A plant's sensory response to external stimuli relies on chemical messengers (hormones). Plant hormones affect all aspects of plant life, from flowering to fruit setting and maturation, and from phototropism to leaf fall. Potentially every cell in a plant can produce plant hormones. They can act in their cell of origin or be transported to other portions of the plant body, with many plant responses involving the synergistic or antagonistic interaction of two or more hormones. In contrast, animal hormones are produced in specific glands and transported to a distant site for action, and they act alone.

Plant hormones are a group of unrelated chemical substances that affect plant morphogenesis. Five major plant hormones are traditionally described: auxins, cytokinins, gibberellins, ethylene, and abscisic acid. In addition, other nutrients and environmental conditions can be characterized as growth factors.

Auxins

The term auxin is derived from the Greek word *auxein*, which means "to grow." **Auxins** are the main hormones responsible for cell elongation in phototropism and gravitropism. They also control the differentiation of meristem into vascular tissue, and promote leaf development and arrangement. While many synthetic auxins are used as herbicides, IAA (a type of auxin) is the only naturally occurring auxin that shows physiological activity. Apical dominance—the inhibition of lateral bud formation—is triggered by auxins produced in the apical meristem. Flowering, fruit setting and ripening, and inhibition of abscission (leaf falling) are other plant responses under the direct or indirect control of auxins. Auxins also act as a relay for the effects of the blue light and red/far-red responses.

Commercial use of auxins is widespread in plant nurseries and for crop production. IAA is used as a rooting hormone to promote growth of adventitious roots on cuttings and detached leaves. Applying synthetic auxins to tomato plants in greenhouses promotes normal fruit development. Outdoor application of auxin promotes synchronization of fruit setting and dropping to coordinate the harvesting season. Fruits such as seedless cucumbers can be induced to set fruit by treating unfertilized plant flowers with auxins.

Cytokinins

The effect of cytokinins was first reported when it was found that adding the liquid endosperm of coconuts to developing plant embryos in culture stimulated their growth. The stimulating growth factor was found to be **cytokinin**, a hormone that promotes cytokinesis (cell division). Almost 200 naturally occurring or synthetic cytokinins are known to date. Cytokinins are most abundant in growing tissues, such as roots, embryos, and fruits, where cell division is occurring. Cytokinins are known to delay senescence in leaf tissues, promote mitosis, and stimulate differentiation of the meristem in shoots and roots. Many effects on plant development are under the influence of cytokinins, either in conjunction with auxin or another hormone. For example, apical dominance seems to result from a balance between auxins that inhibit lateral buds, and cytokinins that promote bushier growth.

Gibberellins

Gibberellins (GAs) are a group of about 125 closely related plant hormones that stimulate shoot elongation, seed germination, and fruit and flower maturation. GAs are synthesized in the root and stem apical meristems, young leaves, and seed embryos. In urban areas, GA antagonists are sometimes applied to trees under power lines to control growth and reduce the frequency of pruning.

GAs break dormancy (a state of inhibited growth and development) in the seeds of plants that require exposure to cold or light to germinate. Abscisic acid is a strong antagonist of GA action. Other effects of GAs include gender expression, seedless fruit development, and the delay of senescence in leaves and fruit. Seedless grapes are obtained through standard breeding methods and contain inconspicuous seeds that fail to develop. Because GAs are produced by the seeds, and because fruit development and stem elongation are under GA control, these varieties of grapes would normally produce small fruit in compact clusters. Maturing grapes are routinely treated with GA to promote larger fruit size, as well as looser bunches (longer stems), which reduces the instance of mildew infection (**Figure 22.16**).

Figure 22.16 In grapes, application of gibberellic acid increases the size of fruit and loosens clustering. (credit: Bob Nichols, USDA)

Abscisic Acid

The plant hormone **abscisic acid** (ABA) was first discovered as the agent that causes the abscission or dropping of cotton bolls. However, more recent studies indicate that ABA plays only a minor role in the abscission process. ABA accumulates as a response to stressful environmental conditions, such as dehydration, cold temperatures, or shortened day lengths. Its activity counters many of the growth-promoting effects of GAs and auxins. ABA inhibits stem elongation and induces dormancy in lateral buds.

ABA induces dormancy in seeds by blocking germination and promoting the synthesis of storage proteins. Plants adapted to temperate climates require a long period of cold temperature before seeds germinate. This mechanism protects young plants from sprouting too early during unseasonably warm weather in winter. As the hormone gradually breaks down over winter, the seed is released from dormancy and germinates when conditions are favorable in spring. Another effect of ABA is to promote the development of winter buds; it mediates the conversion of the apical meristem into a dormant bud. Low soil moisture causes an increase in ABA, which causes stomata to close, reducing water loss in winter buds.

Ethylene

Ethylene is associated with fruit ripening, flower wilting, and leaf fall. Ethylene is unusual because it is a volatile gas (C_2H_4). Hundreds of years ago, when gas street lamps were installed in city streets, trees that grew close to lamp posts developed twisted, thickened trunks and shed their leaves earlier than expected. These effects were caused by ethylene

volatilizing from the lamps.

Aging tissues (especially senescing leaves) and nodes of stems produce ethylene. The best-known effect of the hormone, however, is the promotion of fruit ripening. Ethylene stimulates the conversion of starch and acids to sugars. Some people store unripe fruit, such as avocadoes, in a sealed paper bag to accelerate ripening; the gas released by the first fruit to mature will speed up the maturation of the remaining fruit. Ethylene also triggers leaf and fruit abscission, flower fading and dropping, and promotes germination in some cereals and sprouting of bulbs and potatoes.

Ethylene is widely used in agriculture. Commercial fruit growers control the timing of fruit ripening with application of the gas. Horticulturalists inhibit leaf dropping in ornamental plants by removing ethylene from greenhouses using fans and ventilation.

Plant Responses to Wind and Touch

The shoot of a pea plant winds around a trellis, while a tree grows on an angle in response to strong prevailing winds. These are examples of how plants respond to touch or wind.

The movement of a plant subjected to constant directional pressure is called **thigmotropism**, from the Greek words *thigma* meaning “touch,” and *tropism* implying “direction.” Tendrils are one example of this. The meristematic region of tendrils is very touch sensitive; light touch will evoke a quick coiling response. Cells in contact with a support surface contract, whereas cells on the opposite side of the support expand. Application of jasmonic acid is sufficient to trigger tendril coiling without a mechanical stimulus.

A thigmonastic response is a touch response independent of the direction of stimulus. In the Venus flytrap, two modified leaves are joined at a hinge and lined with thin fork-like tines along the outer edges. Tiny hairs are located inside the trap. When an insect brushes against these trigger hairs, touching two or more of them in succession, the leaves close quickly, trapping the prey. Glands on the leaf surface secrete enzymes that slowly digest the insect. The released nutrients are absorbed by the leaves, which reopen for the next meal.

Thigmomorphogenesis is a slow developmental change in the shape of a plant subjected to continuous mechanical stress. When trees bend in the wind, for example, growth is usually stunted and the trunk thickens. Strengthening tissue, especially xylem, is produced to add stiffness to resist the wind’s force. Researchers hypothesize that mechanical strain induces growth and differentiation to strengthen the tissues. Ethylene and jasmonate are likely involved in thigmomorphogenesis.

22.3 | Reproductive Development and Structure

Introduction

> “Flowers are the sweetest things that God ever made, and forgot to put a soul into.”

Henry Ward Beecher, in *Proverbs from Plymouth Pulpit* (1887)

Flowers have fascinated humans for millenia, with their marvelous shapes and enticing aromas. But the plant has another goal in mind besides pleasing the human eye and nose, and that is reproduction. Sexual reproduction takes place with slight variations in different groups of plants. Plants have two distinct stages in their lifecycle: the gametophyte stage and the sporophyte stage. The haploid **gametophyte** produces the male and female gametes by mitosis in distinct multicellular structures. Fusion of the male and females gametes forms the diploid zygote, which develops into the **sporophyte**. After reaching maturity, the diploid sporophyte produces spores by meiosis, which in turn divide by mitosis to produce the haploid gametophyte. The new gametophyte produces gametes, and the cycle continues. This is the alternation of generations, and is typical of plant reproduction (**Figure 22.17**).

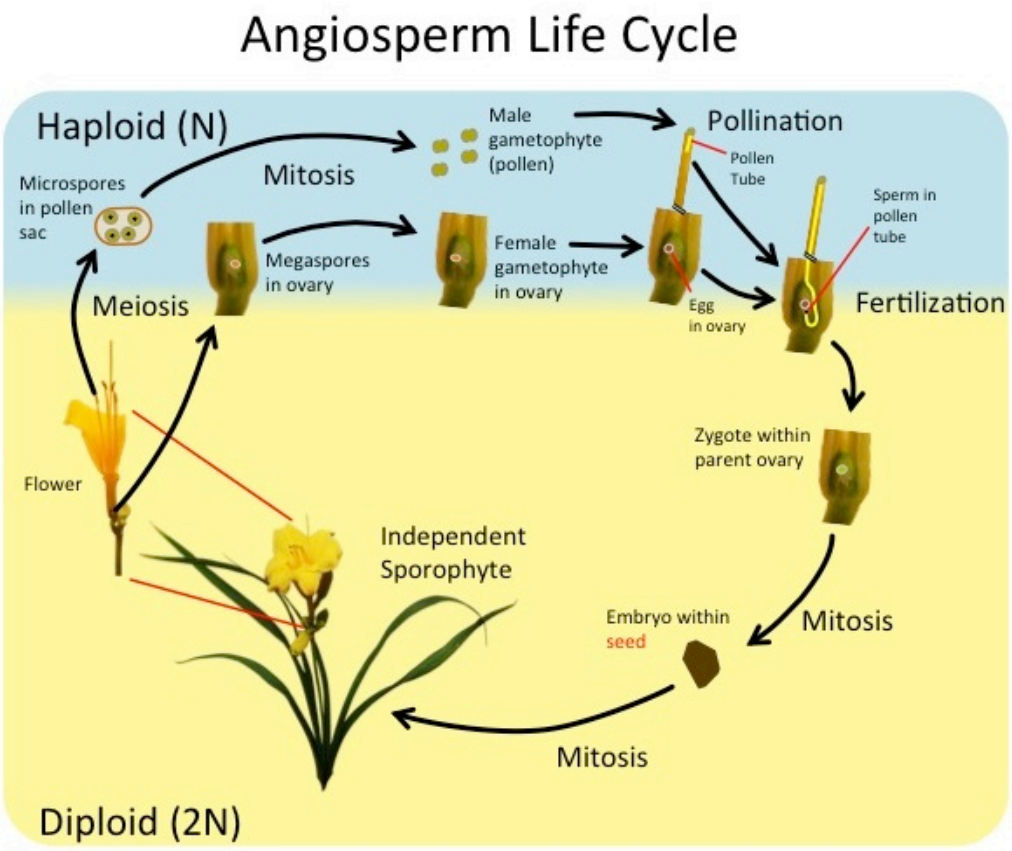


Figure 22.17 The angiosperm life cycle with alternation of generations is depicted in this diagram. Work by Robert A. Bear and Eva Horne

The life cycle of higher plants is dominated by the sporophyte stage, with the gametophyte borne on the sporophyte. In ferns, the gametophyte is free-living and very distinct in structure from the diploid sporophyte. In bryophytes, such as mosses, the haploid gametophyte is more developed than the sporophyte.

During the vegetative phase of growth, plants increase in size and produce a shoot system and a root system. As they enter the reproductive phase, some of the branches start to bear flowers. Many flowers are borne singly, whereas some are borne in clusters. The flower is borne on a stalk known as a receptacle. Flower shape, color, and size are unique to each species, and are often used by taxonomists to classify plants.

Sexual Reproduction in Angiosperms

The lifecycle of angiosperms follows the alternation of generations explained previously. The haploid gametophyte alternates with the diploid sporophyte during the sexual reproduction process of angiosperms. Flowers contain the plant's reproductive structures.

Flower Structure

A typical flower has four main parts—or whorls—known as the calyx, corolla, androecium, and gynoecium (**Figure 22.18**). The outermost whorl of the flower has green, leafy structures known as sepals. The sepals, collectively called the calyx, help to protect the unopened bud. The second whorl is comprised of petals—usually, brightly colored—collectively called the corolla. The number of sepals and petals varies depending on whether the plant is a monocot or eudicot. In monocots, petals usually number three or multiples of three; in eudicots, the number of petals is four or five, or multiples of four and five. Together, the calyx and corolla are known as the perianth. The third whorl contains the male reproductive structures and is known as the androecium. The androecium has stamens with anthers that contain the microsporangia. The innermost group of structures in the flower is the gynoecium, or the female reproductive component(s). The carpel is the individual unit of the gynoecium and has a stigma, style, and ovary. A flower may have one or multiple carpels.

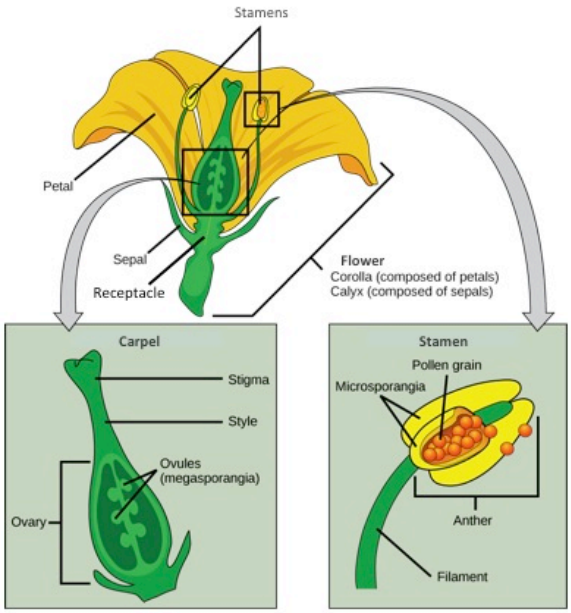


Figure 22.18 The four main parts of the flower are the calyx, corolla, androecium, and gynoecium. The androecium is the sum of all the male reproductive organs, and the gynoecium is the sum of the female reproductive organs. (credit: modification of work by Mariana Ruiz Villareal)

If the anther is missing, what type of reproductive structure will the flower be unable to produce? What term is used to describe an incomplete flower lacking the androecium? What term describes an incomplete flower lacking a gynoecium?

If all four whorls (the calyx, corolla, androecium, and gynoecium) are present, the flower is described as complete. If any of the four parts is missing, the flower is known as incomplete. Flowers that contain both an androecium and a gynoecium are called perfect, androgynous or hermaphrodites. There are two types of incomplete flowers: staminate flowers contain only an androecium, and carpellate flowers have only a gynoecium (**Figure 22.19**).

Figure 22.19 The corn plant has both staminate (male) and carpellate (female) flowers. Staminate flowers, which are clustered in the tassel at the tip of the stem, produce pollen grains. Carpellate flower are clustered in the immature ears. Each strand of silk is a stigma. The corn kernels are seeds that develop on the ear after fertilization. Also shown is the lower stem and root.

If both male and female flowers are borne on the same plant, the species is called monoecious (meaning “one home”): examples are corn and pea. Species with male and female flowers borne on separate plants are termed dioecious, or “two homes,” examples of which are *C. papaya* and *Cannabis*. The ovary, which may contain one or multiple ovules, may be placed above other flower parts, which is referred to as superior; or, it may be placed below the other flower parts, referred to as inferior (**Figure 22.20**).

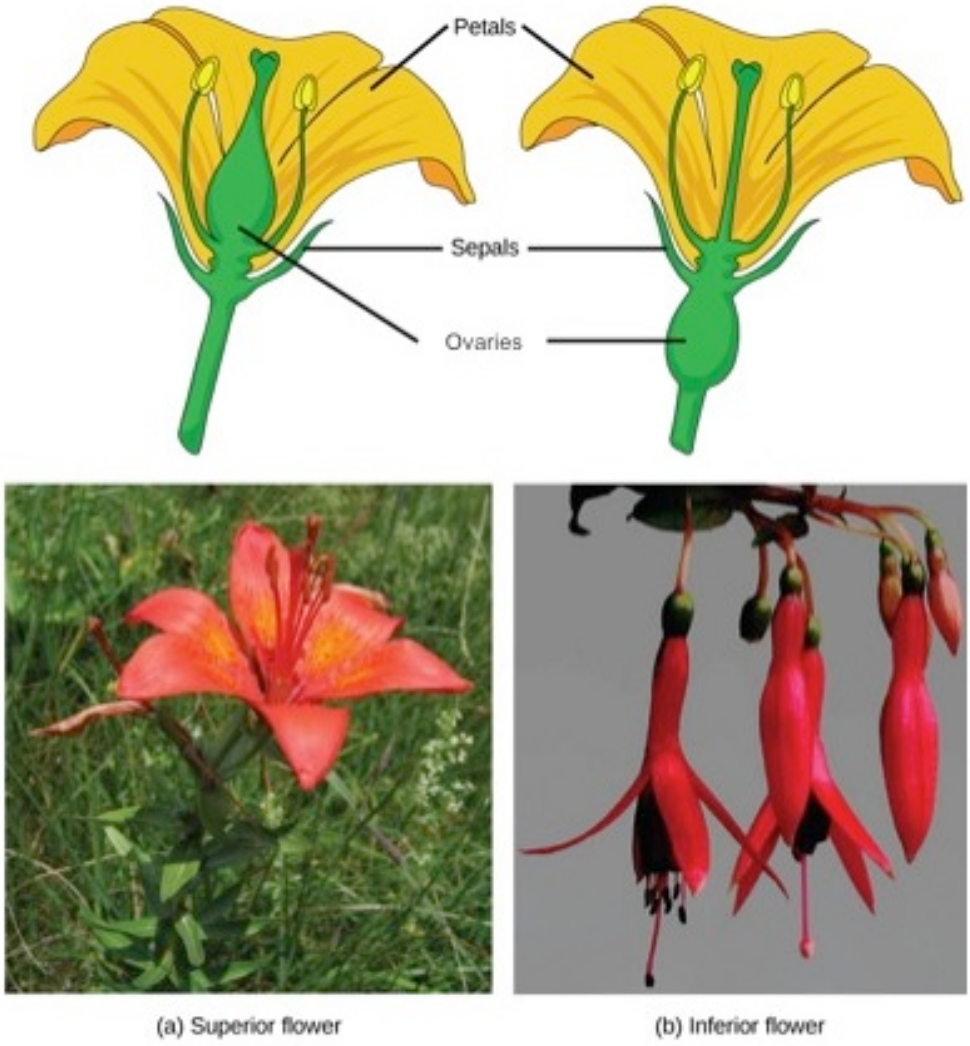


Figure 22.20 The (a) lily is a superior flower, which has the ovary above the other flower parts. (b) Fuchsia is an inferior flower, which has the ovary beneath other flower parts. (credit a photo: modification of work by Benjamin Zwittnig; credit b photo: modification of work by "Koshy Koshy"/Flickr)

Male Gametophyte (The Pollen Grain)

The male gametophyte develops and reaches maturity in an immature anther. In a plant's male reproductive organs, development of pollen takes place in a structure known as the microsporangium (**Figure 22.21**). The microsporangia, which are usually bi-lobed, are pollen sacs in which the microspores develop into pollen grains. These are found in the anther, which is at the end of the stamen—the long filament that supports the anther.

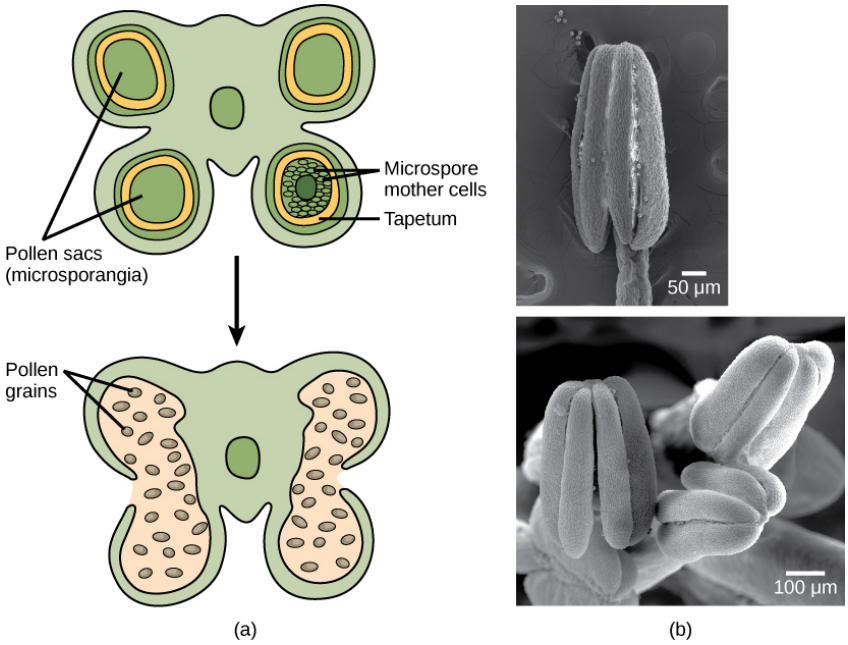


Figure 22.21 Shown is (a) a cross section of an anther at two developmental stages. The immature anther (top) contains four microsporangia, or pollen sacs. Each microsporangium contains hundreds of microspore mother cells that will each give rise to four pollen grains. The tapetum supports the development and maturation of the pollen grains. Upon maturation of the pollen (bottom), the pollen sac walls split open and the pollen grains (male gametophytes) are released. (b) In these scanning electron micrographs, pollen sacs are ready to burst, releasing their grains. (credit b: modification of work by Robert R. Wise; scale-bar data from Matt Russell)

Within the microsporangium, the microspore mother cell divides by meiosis to give rise to four microspores, each of which will ultimately form a pollen grain (**Figure 22.22**). An inner layer of cells, known as the tapetum, provides nutrition to the developing microspores and contributes key components to the pollen wall. Mature pollen grains contain two cells: a generative cell and a pollen tube cell. The generative cell is contained within the larger pollen tube cell. Upon germination, the tube cell forms the pollen tube through which the generative cell migrates to enter the ovary. During its transit inside the pollen tube, the generative cell divides to form two male gametes (sperm cells). Upon maturity, the microsporangia burst, releasing the pollen grains from the anther.

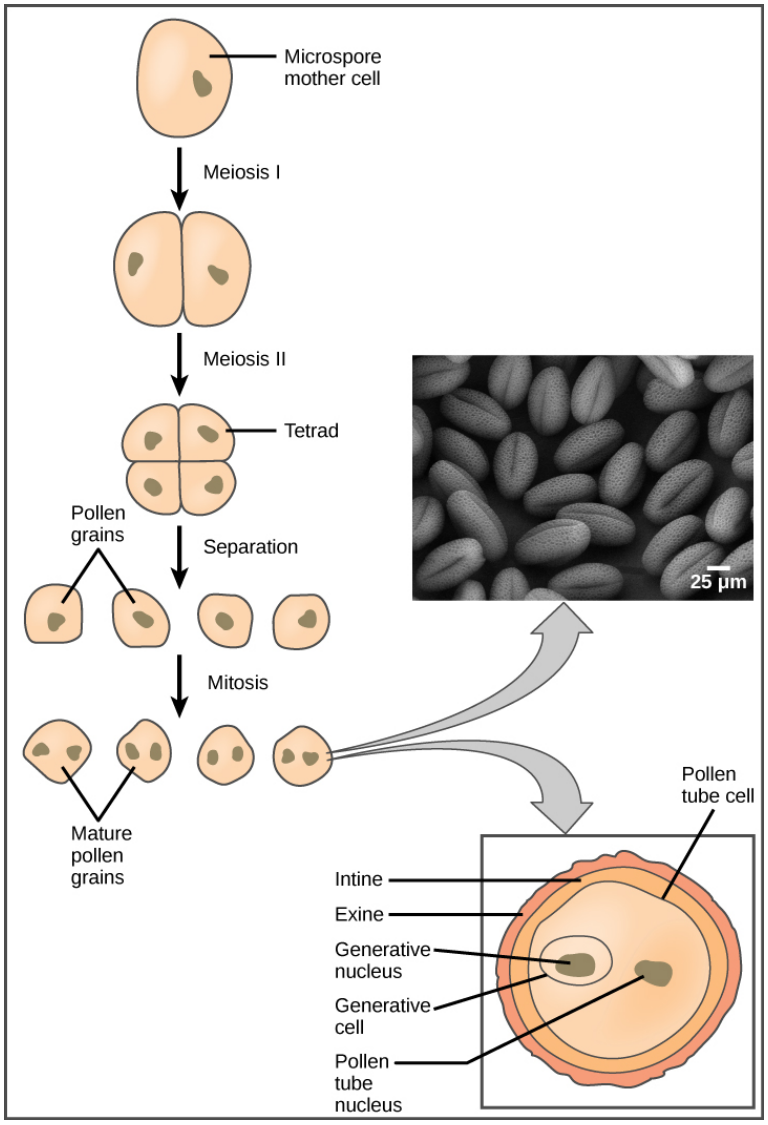


Figure 22.22 Pollen develops from the microspore mother cells. The mature pollen grain is composed of two cells: the pollen tube cell and the generative cell, which is inside the tube cell. The pollen grain has two coverings: an inner layer (intine) and an outer layer (exine). The inset scanning electron micrograph shows *Arabidopsis lyrata* pollen grains. (credit "pollen micrograph": modification of work by Robert R. Wise; scale-bar data from Matt Russell)

Each pollen grain has two coverings: the exine (thicker, outer layer) and the intine (**Figure 22.22**). The exine contains sporopollenin, a complex waterproofing substance supplied by the tapetal cells. Sporopollenin allows the pollen to survive under unfavorable conditions and to be carried by wind, water, or biological agents without undergoing damage.

Female Gametophyte (The Embryo Sac)

While the details may vary between species, the overall development of the female gametophyte has two distinct phases. First, in the process of megasporogenesis, a single cell in the diploid megasporangium—an area of tissue in the ovules—undergoes meiosis to produce four megaspores, only one of which survives. During the second phase, megagametogenesis, the surviving haploid megaspore undergoes mitosis to produce an eight-nucleate, seven-cell female gametophyte, also known as the megagametophyte or embryo sac. Two of the nuclei—the polar nuclei—move to the equator and fuse, forming a single, diploid central cell. This central cell later fuses with a sperm to form the **triploid endosperm**. Three nuclei position themselves on the end of the embryo sac opposite the micropyle and develop into the antipodal cells, which later degenerate. The nucleus closest to the micropyle becomes the female gamete, or egg cell, and the two adjacent nuclei develop into synergid cells (**Figure 22.23**). The synergids help guide the pollen tube for successful fertilization, after which they disintegrate. Once fertilization is complete, the resulting diploid zygote develops into the embryo, and the fertilized ovule forms the other tissues of the seed.

A double-layered integument protects the megasporangium and, later, the embryo sac. The integument will develop into the

seed coat after fertilization and protect the entire seed. The ovule wall will become part of the fruit. The integuments, while protecting the megasporangium, do not enclose it completely, but leave an opening called the micropyle. The micropyle allows the pollen tube to enter the female gametophyte for fertilization.

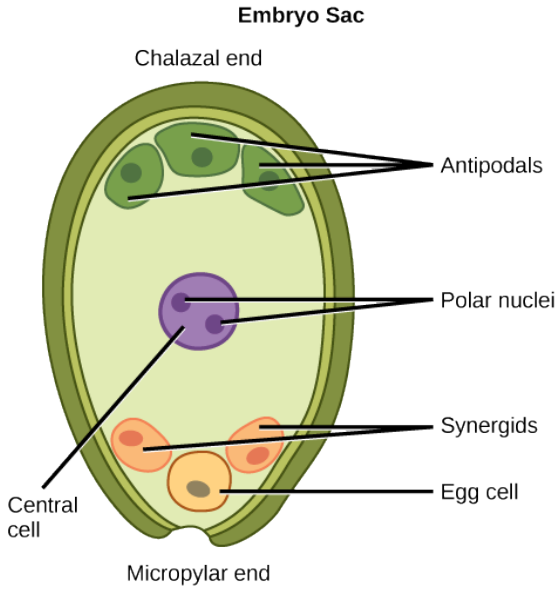


Figure 22.23 As shown in this diagram of the embryo sac in angiosperms, the ovule is covered by integuments and has an opening called a micropyle. Inside the embryo sac are three antipodal cells, two synergids, a central cell, and the egg cell.

22.4 | Pollination and Fertilization

Introduction

> "Nature is full of by-ends. A moth feeds on a petal, in a moment the pollen caught on its breast will be wedding this blossom to another in the next county."

George Iles, in *Canadian Stories* (1918)

In angiosperms, **pollination** is defined as the placement or transfer of pollen from the anther to the stigma of the same flower or another flower. In gymnosperms, pollination involves pollen transfer from the male cone to the female cone. Upon transfer, the pollen germinates to form the pollen tube and the sperm for fertilizing the egg. Pollination has been well studied since the time of Gregor Mendel. Mendel successfully carried out self- as well as cross-pollination in garden peas while studying how characteristics were passed on from one generation to the next. Today's crops are a result of plant breeding, which employs artificial selection to produce the present-day cultivars. A case in point is today's corn, which is a result of years of breeding that started with its ancestor, teosinte. The teosinte that the ancient Mayans originally began cultivating had tiny seeds—vastly different from today's relatively giant ears of corn. Interestingly, though these two plants appear to be entirely different, the genetic difference between them is miniscule.

Pollination takes two forms: self-pollination and cross-pollination. Self-pollination occurs when the pollen from the anther is deposited on the stigma of the same flower, or another flower on the same plant. Cross-pollination is the transfer of pollen from the anther of one flower to the stigma of another flower on a different individual of the same species. Self-pollination occurs in flowers where the stamen and carpel mature at the same time, and are positioned so that the pollen can land on the flower's stigma. This method of pollination does not require an investment from the plant to provide nectar and pollen as food for pollinators.

Living species are adapted to ensure survival of their progeny; those that fail become extinct. Genetic diversity is therefore

required so that in changing environmental or stress conditions, some of the progeny can survive. Self-pollination leads to the production of plants with less genetic diversity, since genetic material from the same plant is used to form both gametes, and eventually, the zygote. In contrast, cross-pollination—or out-crossing—leads to greater genetic diversity because the microgametophyte and megagametophyte are derived from different plants.

Because cross-pollination allows for more genetic diversity, plants have developed many ways to avoid self-pollination. In some species, the pollen and the ovary mature at different times. These flowers make self-pollination nearly impossible. By the time pollen matures and has been shed, the stigma of this flower is mature and can only be pollinated by pollen from another flower. Some flowers have developed physical features that prevent self-pollination. The primrose is one such flower. Primroses have evolved two flower types with differences in anther and stigma length: the pin-eyed flower has anthers positioned at the pollen tube's halfway point, and the thrum-eyed flower's stigma is likewise located at the halfway point. Insects easily cross-pollinate while seeking the nectar at the bottom of the pollen tube. This phenomenon is also known as heterostyly. Many plants, such as cucumber, have male and female flowers located on different parts of the plant, thus making self-pollination difficult. In yet other species, the male and female flowers are borne on different plants (dioecious). All of these are barriers to self-pollination; therefore, the plants depend on pollinators to transfer pollen. The majority of pollinators are biotic agents such as insects (like bees, flies, and butterflies), bats, birds, and other animals. Other plant species are pollinated by abiotic agents, such as wind and water.

Pollination by Insects

Bees are perhaps the most important pollinator of many garden plants and most commercial fruit trees (**Figure 22.24**). The most common species of bees are bumblebees and honeybees. Since bees cannot see the color red, bee-pollinated flowers usually have shades of blue, yellow, or other colors. Bees collect energy-rich pollen or nectar for their survival and energy needs. They visit flowers that are open during the day, are brightly colored, have a strong aroma or scent, and have a tubular shape, typically with the presence of a nectar guide. A nectar guide includes regions on the flower petals that are visible only to bees, and not to humans; it helps to guide bees to the center of the flower, thus making the pollination process more efficient. The pollen sticks to the bees' fuzzy hair, and when the bee visits another flower, some of the pollen is transferred to the second flower. Recently, there have been many reports about the declining population of honeybees. Many flowers will remain unpollinated and not bear seed if honeybees disappear. The impact on commercial fruit growers could be devastating.

Figure 22.24 Insects, such as bees, are important agents of pollination. (credit: photo by D. A. Rintoul)

Many flies are attracted to flowers that have a decaying smell or an odor of rotting flesh. These flowers, which produce nectar, usually have dull colors, such as brown or purple. They are found on the corpse flower or voodoo lily (*Amorphophallus*), dragon arum (*Dracunculus*), and carrion flower (*Stapleia*, *Rafflesia*). The nectar provides energy, whereas the pollen provides protein. Wasps are also important insect pollinators, and pollinate many species of figs.

Butterflies, such as the monarch, pollinate many garden flowers and wildflowers, which usually occur in clusters. These flowers are brightly colored, have a strong fragrance, are open during the day, and have nectar guides to make access to nectar easier. The pollen is picked up and carried on the butterfly's limbs. Moths, on the other hand, pollinate flowers during the late afternoon and night. The flowers pollinated by moths are pale or white and are flat, enabling the moths to land. One well-studied example of a moth-pollinated plant is the yucca plant, which is pollinated by the yucca moth. The shape of the flower and moth have adapted in such a way as to allow successful pollination. The moth deposits pollen on the sticky stigma for fertilization to occur later. The female moth also deposits eggs into the ovary. As the eggs develop into larvae, they obtain food from the flower and developing seeds. Thus, both the insect and flower benefit from each other in this symbiotic relationship. The corn earworm moth and Gaura plant have a similar relationship (**Figure 22.25**).

Figure 22.25 A corn earworm sips nectar from a night-blooming Gaura plant. (credit: Juan Lopez, USDA ARS)

Pollination by Bats

In the tropics and deserts, bats are often the pollinators of nocturnal flowers such as agave, guava, and morning glory. The flowers are usually large and white or pale-colored; thus, they can be distinguished from the dark surroundings at night. The flowers have a strong, fruity, or musky fragrance and produce large amounts of nectar. They are naturally large and wide-mouthed to accommodate the head of the bat. As the bats seek the nectar, their faces and heads become covered with pollen, which is then transferred to the next flower.

Pollination by Birds

Many species of small birds, such as hummingbirds and the Australasian birds known as honeyeaters (**Figure 22.26**), are pollinators for a wide variety of plants. Flowers visited by birds are usually sturdy and are oriented in such a way as to allow the birds to stay near the flower without getting their wings entangled in the nearby flowers. The flower typically has a curved, tubular shape, which allows access for the bird's beak. Brightly colored, odorless flowers that are open during the day are pollinated by birds. As a bird seeks energy-rich nectar, pollen is deposited on the bird's head and neck and is then transferred to the next flower it visits. Botanists have been known to determine the range of extinct plants by collecting and identifying pollen from 200-year-old bird specimens from the same site.

Figure 22.26 Hummingbirds (shown at left) have adaptations (long beaks, ability to hover, etc.) that allow them to reach the nectar of certain tubular flowers, but in the process they can collect pollen on their beaks or heads and then transfer it to another flower. Other birds, like the Tui (at right), a honeyeater found in New Zealand, also collect pollen on their foreheads as they forage on nectar. (credit: photos by D. A. Rintoul)

Pollination by Wind

Most species of conifers, and many angiosperms, such as grasses, maples and oaks, are pollinated by wind. Pine cones are brown and unscented, while the flowers of wind-pollinated angiosperm species are usually green, small, may have small or no petals, and produce large amounts of pollen. Unlike the typical insect-pollinated flowers, flowers adapted to pollination by wind do not produce nectar or scent. In wind-pollinated species, the microsporangia hang out of the flower, and, as the wind blows, the lightweight pollen is carried with it (**Figure 22.27**). The flowers usually emerge early in the spring, before the leaves, so that the leaves do not block the movement of the wind. The pollen is deposited on the exposed feathery stigma of the flower (**Figure 22.28**).

Figure 22.27 A person knocks pollen from a pine tree.

(a) (b)

Figure 22.28 These male (a) and female (b) catkins are from the goat willow tree (*Salix caprea*). Note how both structures are light and feathery to better disperse and catch the wind-blown pollen.

Pollination by Water

Some weeds, such as Australian sea grass and pond weeds, are pollinated by water. The pollen floats on water, and when it comes into contact with the flower, it is deposited inside the flower.

evolution CONNECTION

Pollination by Deception

Orchids are highly valued flowers, with many rare varieties (Figure 22.29). They grow in a range of specific habitats, mainly in the tropics of Asia, South America, and Central America. At least 25,000 species of orchids have been identified.

Figure 22.29 Certain orchids use food deception or sexual deception to attract pollinators. Shown here is a bee orchid (*Ophrys apifera*). (credit: David Evans)

Flowers often attract pollinators with food rewards, in the form of nectar. However, some species of orchid are an exception to this standard: they have evolved different ways to attract the desired pollinators. They use a method known as food deception, in which bright colors and perfumes are offered, but no food. *Anacamptis morio*, commonly known as the green-winged orchid, bears bright purple flowers and emits a strong scent. The bumblebee, its main pollinator, is attracted to the flower because of the strong scent—which usually indicates food for a bee—and in the process, picks up the pollen to be transported to another flower.

Other orchids use sexual deception. *Chiloglottis trapeziformis* emits a compound that smells the same as the pheromone emitted by a female wasp to attract male wasps. The male wasp is attracted to the scent, lands on the orchid flower, and in the process, transfers pollen. Some orchids, like the Australian hammer orchid, use scent as well as visual trickery in yet another sexual deception strategy to attract wasps. The flower of this orchid mimics the appearance of a female wasp and emits a pheromone. The male wasp tries to mate with what appears to be a female wasp, and in the process, picks up pollen, which it then transfers to the next counterfeit mate.

Double Fertilization

After pollen is deposited on the stigma, it must germinate and grow through the style to reach the ovule. The microspores, or the pollen, contain two cells: the pollen tube cell and the generative cell. The pollen tube cell grows into a pollen tube through which the generative cell travels. The germination of the pollen tube requires water, oxygen, and certain chemical signals. As it travels through the style to reach the embryo sac, the pollen tube's growth is supported by the tissues of the style. In the meantime, if the generative cell has not already split into two cells, it now divides to form two sperm cells. The pollen tube is guided by the chemicals secreted by the synergids present in the embryo sac, and it enters the ovule sac through the micropyle. Of the two sperm cells, one sperm fertilizes the egg cell, forming a diploid zygote; the other sperm fuses with the two polar nuclei, forming a triploid cell that develops into the **endosperm**. Together, these two fertilization events in angiosperms are known as **double fertilization** (Figure 22.30). After fertilization is complete, no other sperm can

enter. The fertilized ovule forms the seed, whereas the tissues of the ovary become the fruit, usually enveloping the seed.

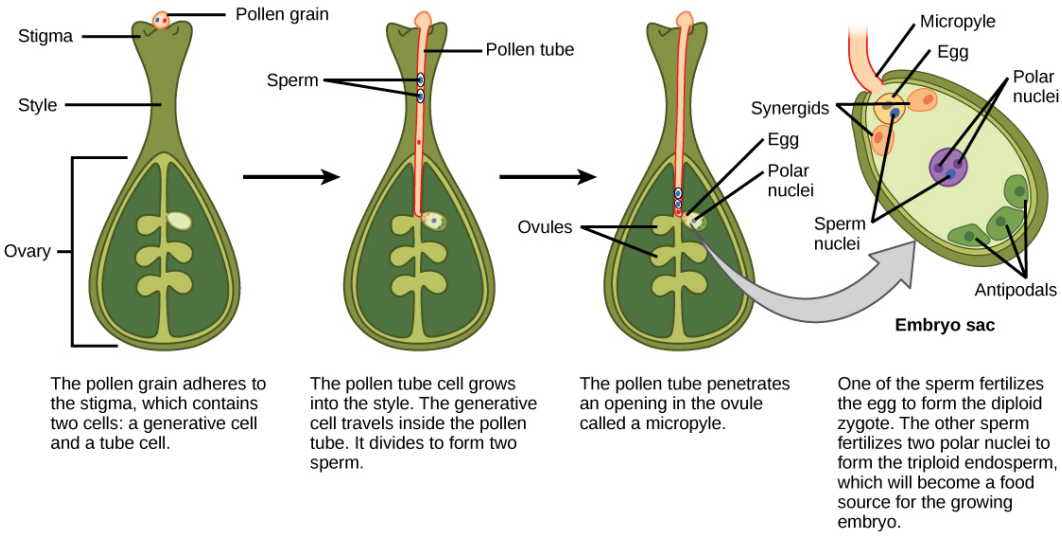


Figure 22.30 In angiosperms, one sperm fertilizes the egg to form the 2*n* zygote, and the other sperm fertilizes the central cell to form the 3*n* endosperm. This is called a double fertilization.

After fertilization, the zygote divides to form two cells: the upper cell, or terminal cell, and the lower, or basal, cell. The division of the basal cell gives rise to the suspensor, which eventually makes connection with the maternal tissue. The suspensor provides a route for nutrition to be transported from the mother plant to the growing embryo. The terminal cell also divides, giving rise to a globular-shaped proembryo (**Figure 22.31a**). In eudicots, the developing embryo has a heart shape, due to the presence of the two rudimentary **cotyledons** (**Figure 22.31b**). In non-endospermic eudicots, such as *Capsella bursa*, the endosperm develops initially, but is then digested, and the food reserves are moved into the two cotyledons. As the embryo and cotyledons enlarge, they run out of room inside the developing seed, and are forced to bend (**Figure 22.31c**). Ultimately, the embryo and cotyledons fill the seed (**Figure 22.31d**), and the seed is ready for dispersal. Embryonic development is suspended after some time, and growth is resumed only when the seed germinates. The developing seedling will rely on the food reserves stored in the cotyledons until the first set of leaves begin photosynthesis.

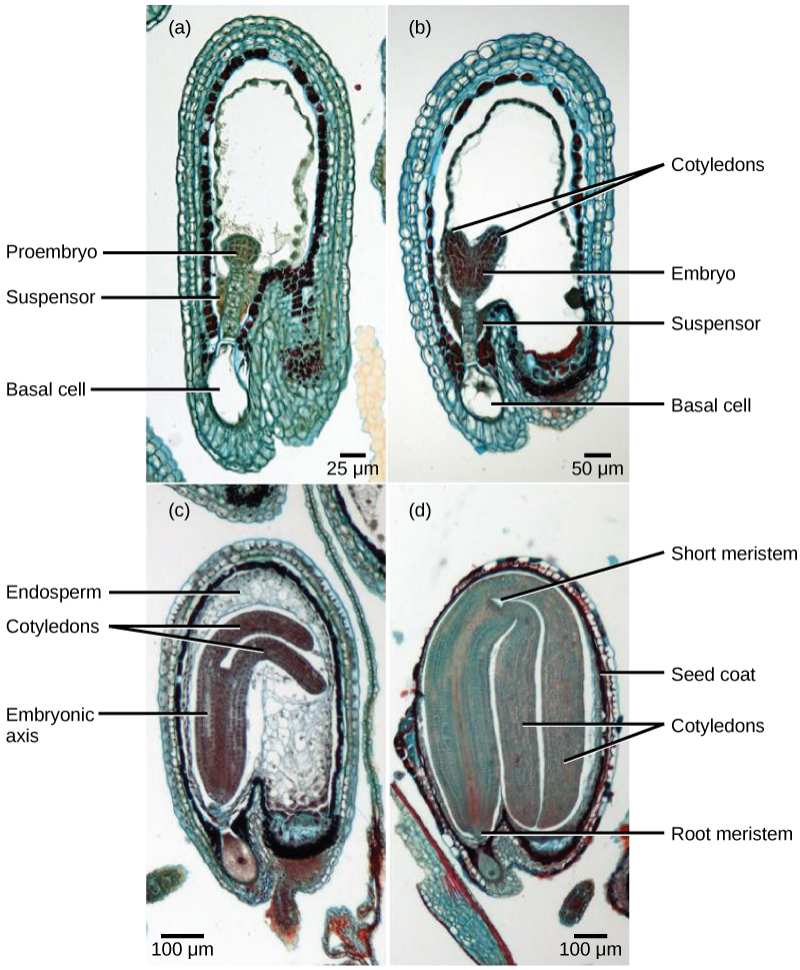


Figure 22.31 Shown are the stages of embryo development in the ovule of a shepherd's purse (*Capsella bursa*). After fertilization, the zygote divides to form an upper terminal cell and a lower basal cell. (a) In the first stage of development, the terminal cell divides, forming a globular pro-embryo. The basal cell also divides, giving rise to the suspensor. (b) In the second stage, the developing embryo has a heart shape due to the presence of cotyledons. (c) In the third stage, the growing embryo runs out of room and starts to bend. (d) Eventually, it completely fills the seed. (credit: modification of work by Robert R. Wise; scale-bar data from Matt Russell)

Development of the Seed

The mature ovule develops into the seed. A typical seed contains a seed coat, cotyledons, endosperm, and a single embryo (**Figure 22.32**).

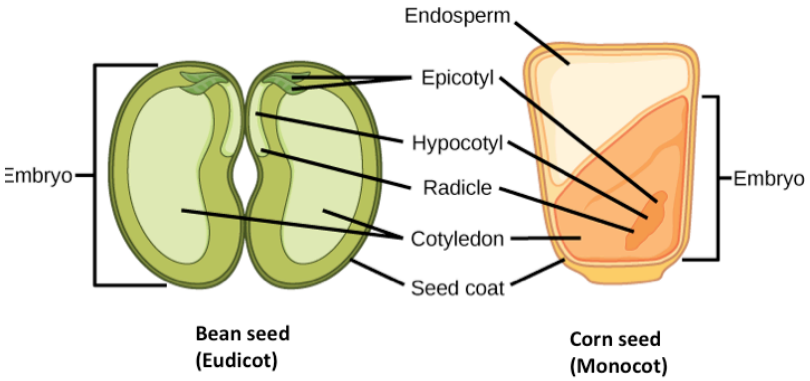


Figure 22.32 The structures of eudicot and monocot seeds are shown. Eudicots (left) have two cotyledons. Monocots, such as corn (right), have one cotyledon, called the scutellum; it channels nutrition to the growing embryo. Both monocot and eudicot embryos have a plumule that forms the leaves, a hypocotyl that forms the stem, and a radicle that forms the root. The embryonic axis comprises everything between the plumule and the radicle, not including the cotyledon(s).

The storage of food reserves in angiosperm seeds differs between monocots and eudicots. In monocots, such as corn and wheat, the single cotyledon is connected directly to the embryo via vascular tissue (xylem and phloem). Food reserves are stored in the large endosperm. Upon germination, enzymes are secreted by the aleurone, a single layer of cells just inside the seed coat that surrounds the endosperm and embryo. The enzymes degrade the stored carbohydrates, proteins and lipids, the products of which are absorbed by the cotyledon and transported to the developing embryo. Therefore, the cotyledon can be seen to be an absorptive organ, not a storage organ.

The two cotyledons in the eudicot seed also have vascular connections to the embryo. In endospermic eudicots, the food reserves are stored in the endosperm. During germination, the two cotyledons therefore act as absorptive organs to take up the enzymatically released food reserves, much like in monocots (monocots, by definition, also have endospermic seeds). Tobacco (*Nicotiana tabaccum*), tomato (*Solanum lycopersicum*), and pepper (*Capsicum annuum*) are examples of endospermic eudicots. In non-endospermic eudicots, the triploid endosperm develops normally following double fertilization, but the endosperm food reserves are quickly remobilized and moved into the developing cotyledon for storage. The two halves of a peanut seed (*Arachis hypogaea*) and the split peas (*Pisum sativum*) of split pea soup are individual cotyledons loaded with food reserves.

The seed, along with the ovule, is protected by a seed coat that is formed from the integuments of the ovule sac. The embryo consists of three parts: the plumule, the radicle, and the hypocotyl. The portion of the embryo between the cotyledon attachment point and the radicle is known as the **hypocotyl** (hypocotyl means “below the cotyledons”). The embryo terminates in a **radicle** (the embryonic root), which is the region from which the root will develop. In eudicots, the hypocotyls extend above ground, giving rise to the stem of the plant. In monocots, the hypocotyl does not show above ground because monocots do not exhibit stem elongation. The part of the embryonic axis that projects above the cotyledons is known as the **epicotyl**. The plumule is composed of the epicotyl, young leaves, and the shoot apical meristem.

Upon germination in eudicot seeds (**Figure 22.33**), the epicotyl is shaped like a hook with the plumule pointing downwards. This shape is called the plumule hook, and it persists as long as germination proceeds in the dark. Therefore, as the epicotyl pushes through the tough and abrasive soil, the plumule is protected from damage. Upon exposure to light, the hypocotyl hook straightens out, the young foliage leaves face the sun and expand, and the epicotyl continues to elongate. During this time, the radicle is also growing and producing the primary root. As it grows downward to form the tap root, lateral roots branch off to all sides, producing the typical eudicot tap root system.

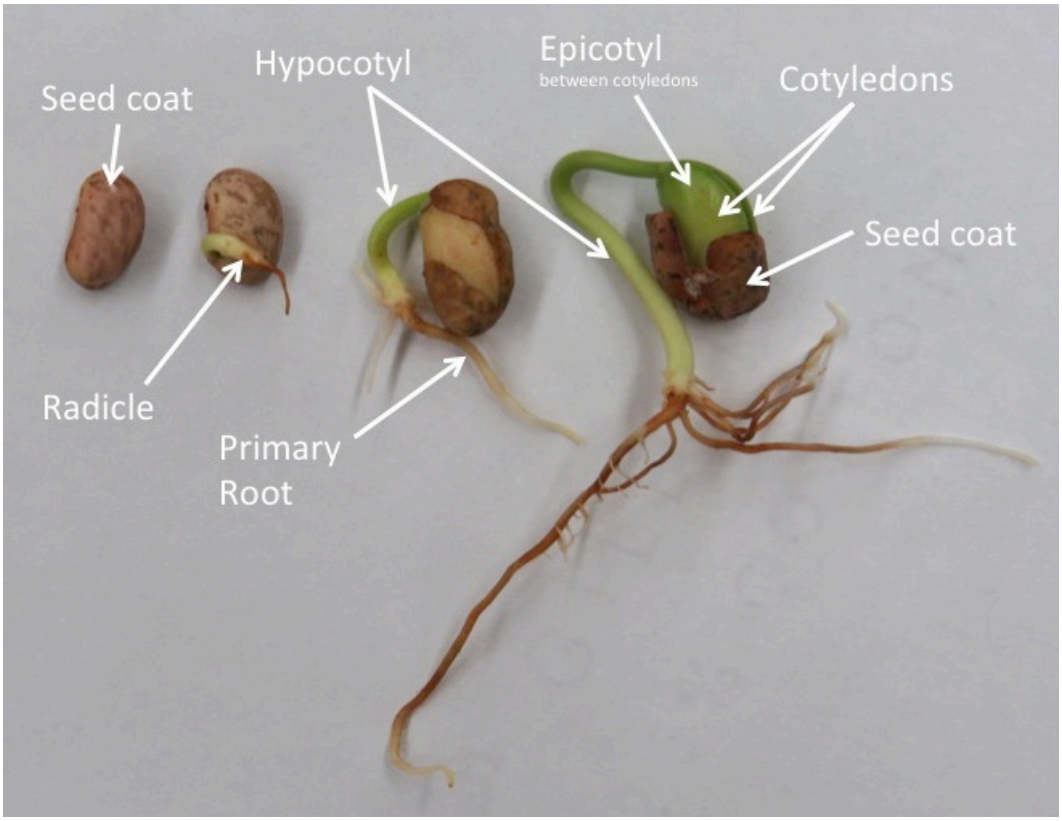


Figure 22.33 As this eudicot bean seed germinates, the primary root, or radicle, emerges first, followed by the hypocotyl and cotyledons. Work by Robert A. Bear

As a monocot seed germinates (**Figure 22.34**), the primary root emerges, protected by the root-tip covering: the coleorhiza. Next, the primary shoot emerges, protected by the coleoptile: the covering of the shoot tip. Upon exposure to light (i.e. when the plumule has exited the soil and the protective coleoptile is no longer needed), elongation of the coleoptile ceases and the leaves expand and unfold. At the other end of the embryon, the primary root soon dies, while other, adventitious roots (roots that do not arise from the usual place – i.e. the root) emerge from the base of the stem. This gives the monocot a fibrous root system.

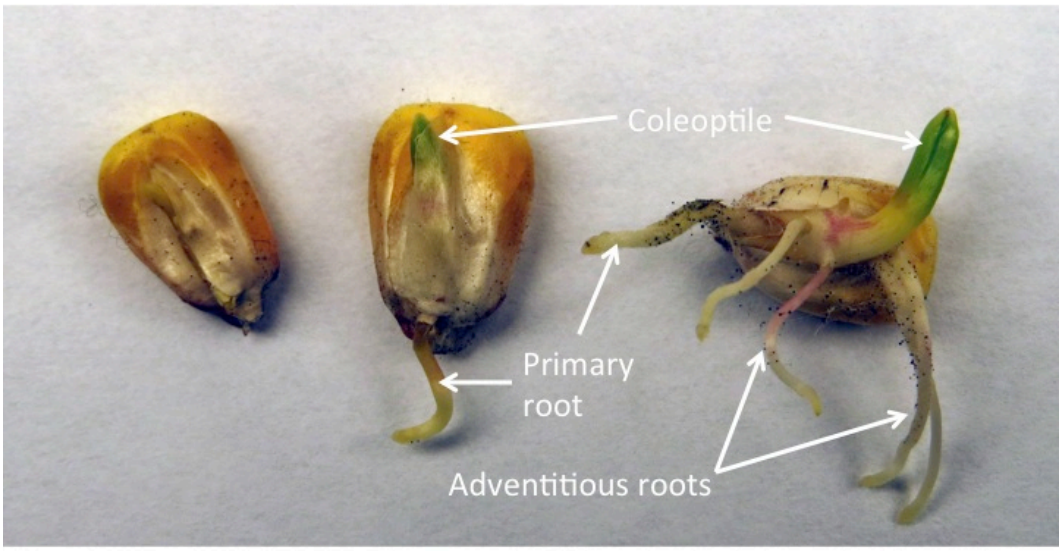


Figure 22.34 As this monocot corn seed germinates, the primary root, or radicle, emerges first, followed by the primary shoot, or coleoptile, and the adventitious roots. Work by Robert A. Bear

Seed Germination

Many mature seeds enter a period of inactivity, or extremely low metabolic activity: a process known as dormancy, which may last for months, years or even centuries. Dormancy helps keep seeds viable during unfavorable conditions. Upon a return to favorable conditions, seed germination takes place. Favorable conditions could be as diverse as moisture, light, cold, fire, or chemical treatments. After heavy rains, many new seedlings emerge. Forest fires also lead to the emergence of new seedlings. Some seeds require vernalization (cold treatment) before they can germinate. This guarantees that seeds produced by plants in temperate climates will not germinate until the spring. Plants growing in hot climates may have seeds that need a heat treatment in order to germinate, to avoid germination in the hot, dry summers. In many seeds, the presence of a thick seed coat retards the ability to germinate. Scarification, which includes mechanical or chemical processes to soften the seed coat, is often employed before germination. Presoaking in hot water, or passing through an acid environment, such as an animal's digestive tract, may also be employed.

Depending on seed size, the time taken for a seedling to emerge may vary. Species with large seeds have enough food reserves to germinate deep below ground, and still extend their epicotyl all the way to the soil surface. Seeds of small-seeded species usually require light as a germination cue. This ensures the seeds only germinate at or near the soil surface (where the light is greatest). If they were to germinate too far underneath the surface, the developing seedling would not have enough food reserves to reach the sunlight.

Development of Fruit and Fruit Types

After fertilization, the ovary of the flower usually develops into the fruit. Fruits are usually associated with having a sweet taste; however, not all fruits are sweet. Botanically, the term "fruit" is used for a ripened ovary. In most cases, flowers in which fertilization has taken place will develop into fruits, and flowers in which fertilization has not taken place will not. Some fruits develop from the ovary and are known as true fruits, whereas others develop from other parts of the female gametophyte and are known as accessory fruits. The fruit encloses the seeds and the developing embryo, thereby providing it with protection. Fruits are of many types, depending on their origin and texture. The sweet tissue of the blackberry, the red flesh of the tomato, the shell of the peanut, and the hull of corn (the tough, thin part that gets stuck in your teeth when you eat popcorn) are all fruits. As the fruit matures, the seeds also mature.

Fruits may be classified as simple, aggregate, multiple, or accessory, depending on their origin (**Figure 22.35**). If the fruit develops from a single carpel or fused carpels of a single ovary, it is known as a simple fruit, as seen in nuts and beans.

An aggregate fruit is one that develops from more than one carpel, but all are in the same flower: the mature carpels fuse together to form the entire fruit, as seen in the raspberry. Multiple fruit develops from an inflorescence or a cluster of flowers. An example is the pineapple, where the flowers fuse together to form the fruit. Accessory fruits (sometimes called false fruits) are not derived from the ovary, but from another part of the flower, such as the receptacle (strawberry) or the hypanthium (apples and pears).

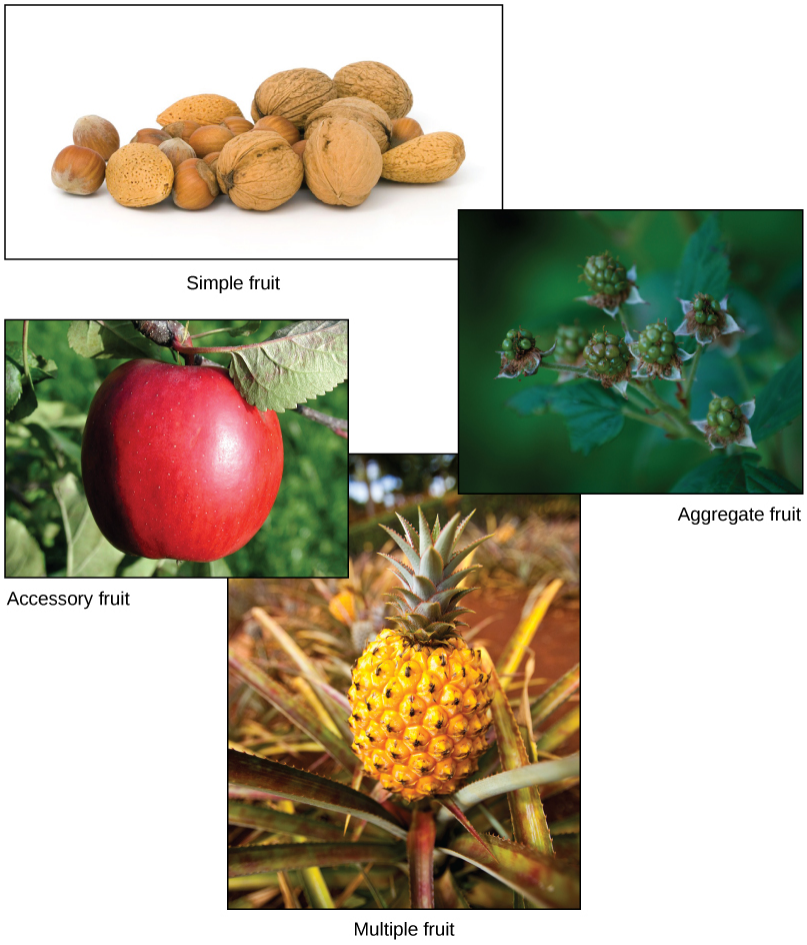


Figure 22.35 There are four main types of fruits. Simple fruits, such as these nuts, are derived from a single ovary. Aggregate fruits, like raspberries, form from many carpels that fuse together. Multiple fruits, such as pineapple, form from a cluster of flowers called an inflorescence. Accessory fruit, like the apple, are formed from a part of the plant other than the ovary. (credit "nuts": modification of work by Petr Kratochvil; credit "raspberries": modification of work by Cory Zanker; credit "pineapple": modification of work by Howie Le; credit "apple": modification of work by Paolo Neo)

Fruits generally have three parts: the exocarp (the outermost skin or covering), the mesocarp (middle part of the fruit), and the endocarp (the inner part of the fruit). Together, all three are known as the pericarp. The mesocarp is usually the fleshy, edible part of the fruit; however, in some fruits, such as the almond, the endocarp is the edible part. In many fruits, two or all three of the layers are fused, and are indistinguishable at maturity. Fruits can be dry or fleshy. Furthermore, fruits can be divided into dehiscent or indehiscent types. Dehiscent fruits, such as peas, readily release their seeds, while indehiscent fruits, like peaches, rely on decay to release their seeds.

Fruit and Seed Dispersal

The fruit has a single purpose: seed dispersal. Seeds contained within fruits need to be dispersed far from the mother plant, so they may find favorable and less competitive conditions in which to germinate and grow.

Some fruit have built-in mechanisms so they can disperse by themselves, whereas others require the help of agents like wind, water, and animals (**Figure 22.36**). Modifications in seed structure, composition, and size help in dispersal. Wind-dispersed fruit are lightweight and may have wing-like appendages that allow them to be carried by the wind. Some have a parachute-like structure to keep them afloat. Some fruits—for example, the dandelion—have hairy, nearly weightless structures that are suited to dispersal by wind.

Seeds dispersed by water are contained in light and buoyant fruit, giving them the ability to float. Coconuts are well known for their ability to float on water to reach land where they can germinate. Similarly, willow and silver birches produce lightweight fruit that can float on water.

Animals and birds eat fruits, and the seeds that are not digested are excreted in their droppings some distance away. Some animals, like squirrels, bury seed-containing fruits for later use; if the squirrel does not find its stash of fruit, and if conditions are favorable, the seeds germinate. Some fruits, like the cocklebur, have hooks or sticky structures that stick to an animal's coat and are then transported to another place. Humans also play a big role in dispersing seeds when they carry fruits to new places and throw away the inedible part that contains the seeds.

All of the above mechanisms allow for seeds to be dispersed through space, much like an animal’s offspring can move to a new location. Seed dormancy, which was described earlier, allows plants to disperse their progeny through time: something animals cannot do. Dormant seeds can wait months, years, or even decades for the proper conditions for germination and propagation of the species.

(a) (b) (c)

Figure 22.36 Fruits and seeds are dispersed by various means. (a) Dandelion seeds are dispersed by wind, the (b) coconut seed is dispersed by water, and the (c) acorn is dispersed by animals that cache and then forget it. (credit a: modification of work by "Rosendahl"/Flickr; credit b: modification of work by Shine Oa; credit c: modification of work by Paolo Neo)

23 | INTERACTIONS OF PLANTS WITH THEIR ENVIRONMENT

23.1 | Root and Leaf Structure

Introduction

> “A tree nowhere offers a straight line or a regular curve, but who doubts that root, trunk, boughs, and leaves embody geometry?”

George Iles, in *Canadian Stories* (1918)

We all recognize roots, trunks and leaves as parts of a plant. In this chapter you will learn a bit more about those structures, and how they are parts of plant organ systems.

Plant Organ Systems

In plants, just as in animals, similar cells working together form a tissue. When different types of tissues work together to perform a unique function, they form an organ; organs working together form organ systems. Vascular plants have two distinct organ systems: a shoot system, and a root system. The **shoot system** consists of two portions: the vegetative (non-reproductive) parts of the plant, such as the leaves and the stems, and the reproductive parts of the plant, which include flowers and fruits. The shoot system generally grows above ground, where it absorbs the light needed for photosynthesis. The **root system**, which supports the plants and absorbs water and minerals, is usually underground. Figure 23.1 shows the organ systems of a typical plant.

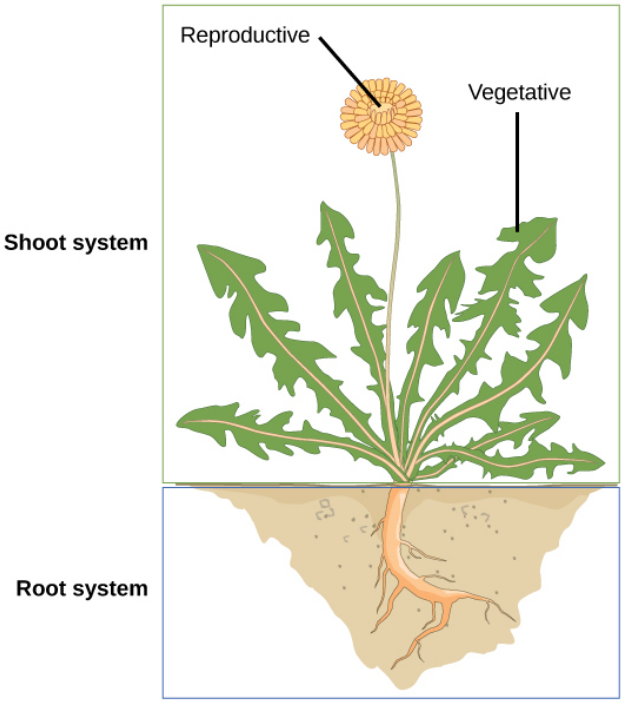


Figure 23.1 The shoot system of a plant consists of leaves, stems, flowers, and fruits. The root system anchors the plant while absorbing water and minerals from the soil.

Roots

The roots of seed plants have three major functions: anchoring the plant to the soil, absorbing water and minerals and transporting them upwards, and storing the products of photosynthesis. Some roots are modified to absorb moisture and exchange gases. Most roots are underground. Some plants, however, also have adventitious roots, which emerge above the ground from the shoot.

Types of Root Systems

Root systems are mainly of two types (**Figure 23.2**). Eudicots have a tap root system, while monocots have a fibrous root system. A **tap root system** has a main root that grows down vertically, and from which many smaller lateral roots arise. Dandelions are a good example; their tap roots usually break off when trying to pull these weeds, and they can regrow another shoot from the remaining root). A tap root system penetrates deep into the soil. In contrast, a **fibrous root system** is located closer to the soil surface, and forms a dense network of roots that also helps prevent soil erosion (lawn grasses are a good example, as are wheat, rice, and corn). Some plants have a combination of tap roots and fibrous roots. Plants that grow in dry areas often have deep root systems, whereas plants growing in areas with abundant water are likely to have shallower root systems.

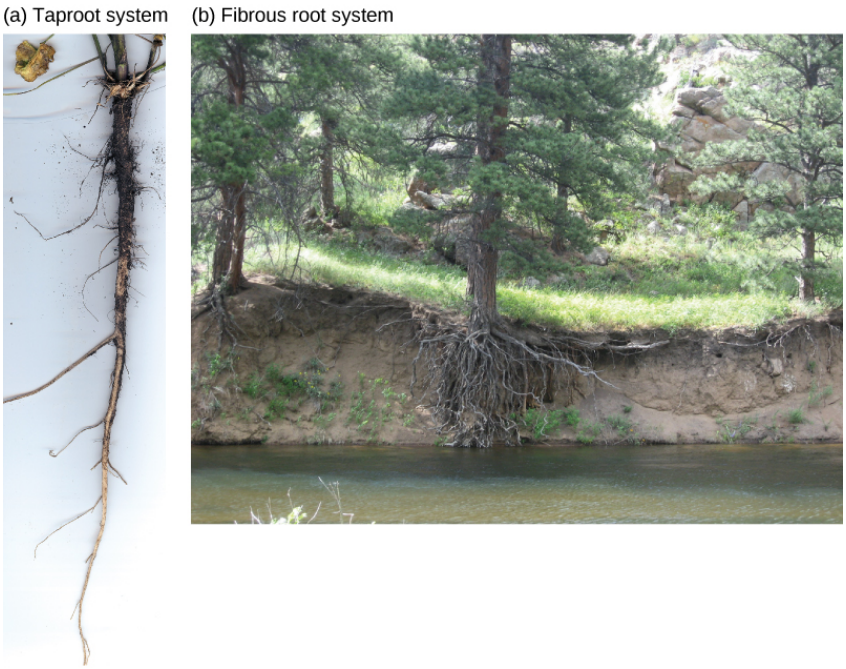


Figure 23.2 (a) Tap root systems have a main root that grows down, while (b) fibrous root systems consist of many small roots. (credit b: modification of work by “Austen Squarepants”/Flickr)

Root Growth and Anatomy

Root growth begins with seed germination. When the plant embryo emerges from the seed, the radicle of the embryo forms the root system. The tip of the root is protected by the **root cap**, a structure exclusive to roots and unlike any other plant structure. The root cap is continuously replaced because it gets damaged easily as the root pushes through soil. The root tip can be divided into three zones: a zone of cell division, a zone of elongation, and a zone of maturation and differentiation (**Figure 23.3**). The zone of cell division is closest to the root tip; it is made up of the actively dividing cells of the root meristem. The zone of elongation is where the newly formed cells increase in length, thereby lengthening the root. Beginning at the first root hair is the zone of cell maturation where the root cells begin to differentiate into special cell types. All three zones are in the first centimeter or so of the root tip.

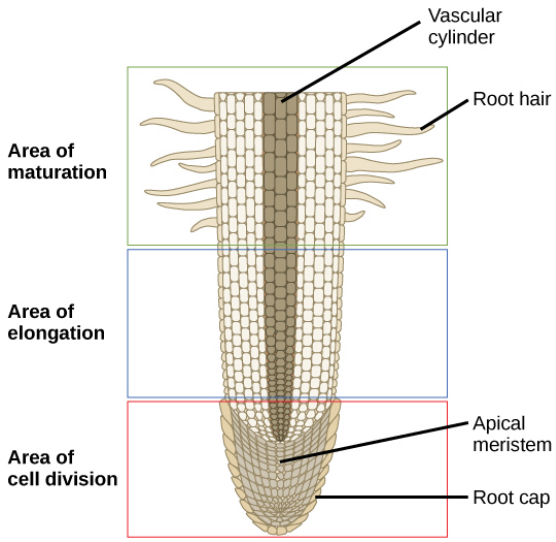


Figure 23.3 A longitudinal view of the root reveals the zones of cell division, elongation, and maturation. Cell division occurs in the apical meristem.

The root has an outer layer of cells called the epidermis, which surrounds areas of ground tissue and vascular tissue. The epidermis provides protection and helps in absorption. **Root hairs**, which are extensions of root epidermal cells, increase the surface area of the root, greatly contributing to the absorption of water and minerals.

Inside the root, the ground tissue forms two regions: the cortex and the pith (**Figure 23.4**). Compared to stems, roots have lots of cortex and little pith. Both regions include cells that store photosynthetic products. The cortex is between the epidermis and the vascular tissue, whereas the pith lies between the vascular tissue and the center of the root.

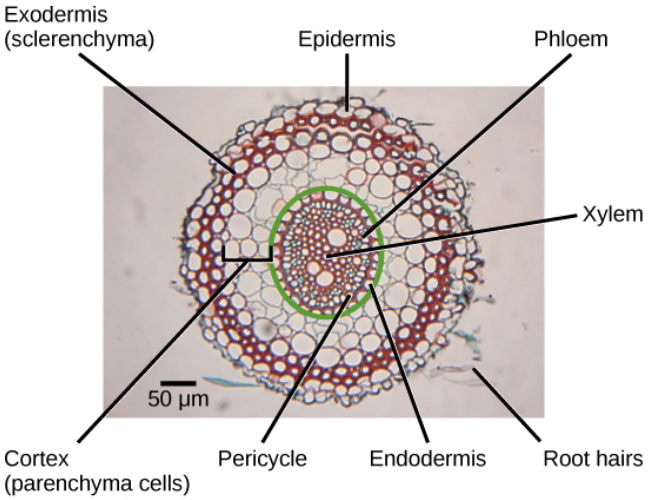


Figure 23.4 Staining reveals different cell types in this light micrograph of a wheat (*Triticum*) root cross section. Sclerenchyma cells of the exodermis and xylem cells stain red, and phloem cells stain blue. Other cell types stain black. The stele, or vascular tissue, is the area inside endodermis (indicated by a green ring). Root hairs are visible outside the epidermis. (credit: scale-bar data from Matt Russell)

The vascular tissue in the root is arranged in the inner portion of the root, which is called the vascular cylinder (**Figure 23.5**). A layer of cells known as the **endodermis** separates the vascular tissue from the ground tissue in the outer portion of the root. The endodermis is exclusive to roots, and serves as a checkpoint for materials entering the root's vascular system. A waxy substance called suberin is present on the walls of the endodermal cells. This waxy region, known as the **Casparian strip**, forces water and solutes to cross the plasma membranes of endodermal cells instead of slipping between the cells. This ensures that only materials required by the root pass through the endodermis, while toxic substances and pathogens are generally excluded. The outermost cell layer of the root's vascular tissue is the **pericycle**, an area that can give rise to lateral roots. In eudicot roots, the xylem and phloem are arranged alternately in an X shape, whereas in monocot roots, the vascular tissue is arranged in a ring around the pith.

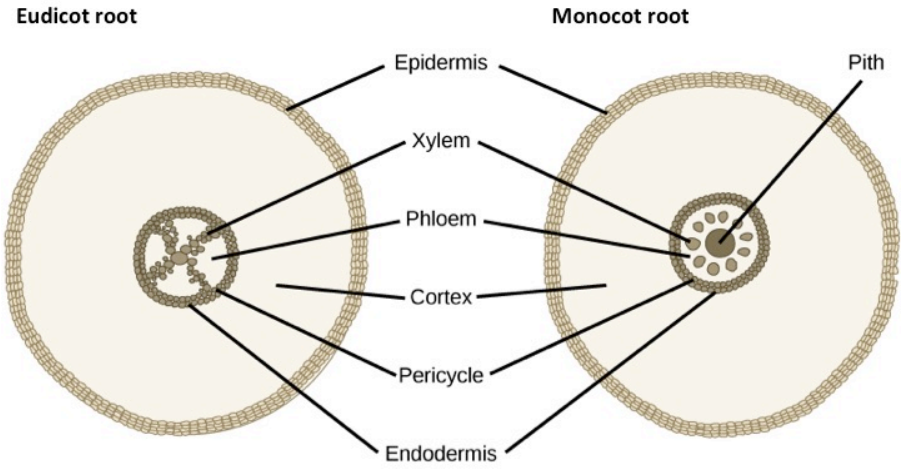


Figure 23.5 In (left) typical eudicots, the vascular tissue forms an X shape in the center of the root. In (right) typical monocots, the phloem cells and the larger xylem cells form a characteristic ring around the central pith.

Root Modifications

Root structures may be modified for specific purposes. For example, some roots are bulbous and store starch. Aerial roots and prop roots are two forms of aboveground roots that provide additional support to anchor the plant. Tap roots, such as carrots, turnips, and beets, are examples of roots that are modified for food storage (**Figure 23.6**).

Figure 23.6 Many vegetables are modified roots.

Epiphytic roots enable a plant to grow on another plant. For example, the epiphytic roots of orchids develop a spongy tissue to absorb moisture. The banyan tree (*Ficus* sp.) begins as an epiphyte, germinating in the branches of a host tree; aerial roots develop from the branches and eventually reach the ground, providing additional support (**Figure 23.7**). In screwpine (*Pandanus* sp.), a palm-like tree that grows in sandy tropical soils, aboveground prop roots develop from the nodes to provide additional support.

(a) (b)

Figure 23.7 The (a) banyan tree, also known as the strangler fig, begins life as an epiphyte in a host tree. Aerial roots extend to the ground and support the growing plant, which eventually strangles the host tree. The (b) screwpine develops aboveground roots that help support the plant in sandy soils. (credit a: modification of work by "psyberartist"/Flickr; credit b: modification of work by David Eikhoff)

Leaves

Leaves are the main sites for photosynthesis: the process by which plants synthesize food. Most leaves are usually green, due to the presence of chlorophyll in the leaf cells. However, some leaves may have different colors, caused by other plant pigments that mask the green chlorophyll.

The thickness, shape, and size of leaves are adapted to the environment. Each variation helps a plant species maximize its chances of survival in a particular habitat. Usually, the leaves of plants growing in tropical rainforests have larger surface areas than those of plants growing in deserts or very cold conditions, which are likely to have a smaller surface area to minimize water loss.

Structure of a Typical Leaf

Each leaf typically has a leaf blade called the lamina, which is also the widest part of the leaf. Some leaves are attached to the plant stem by a petiole. Leaves that do not have a petiole and are directly attached to the plant stem are called sessile leaves. Small green appendages usually found at the base of the petiole are known as stipules. Most leaves have a midrib, which travels the length of the leaf and branches to each side to produce veins of vascular tissue. The edge of the leaf is called the margin. **Figure 23.8** shows the structure of a typical eudicot leaf.

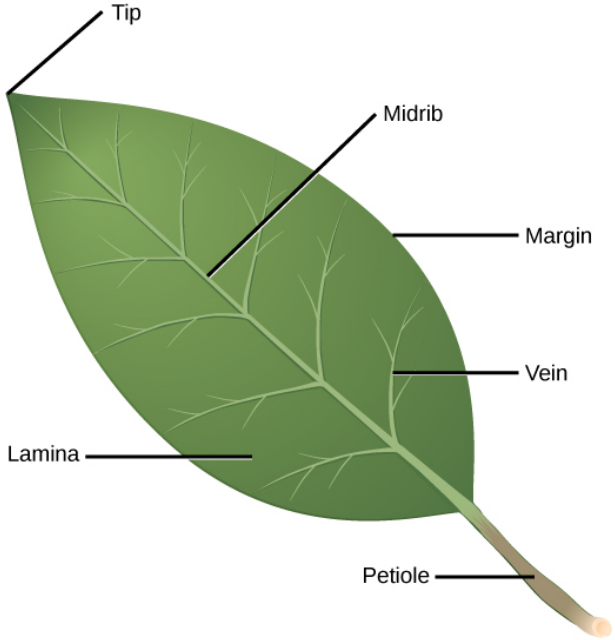


Figure 23.8 Deceptively simple in appearance, a leaf is a highly efficient structure.

Within each leaf, the vascular tissue forms veins. The arrangement of veins in a leaf is called the venation pattern. Monocots

and eudicots differ in their patterns of venation (**Figure 23.9**). Monocots have parallel venation; the veins run in straight lines across the length of the leaf without converging at a point. In eudicots, however, the veins of the leaf have a net-like appearance, forming a pattern known as reticulate venation. One extant plant, the *Ginkgo biloba*, has dichotomous venation where the veins fork.

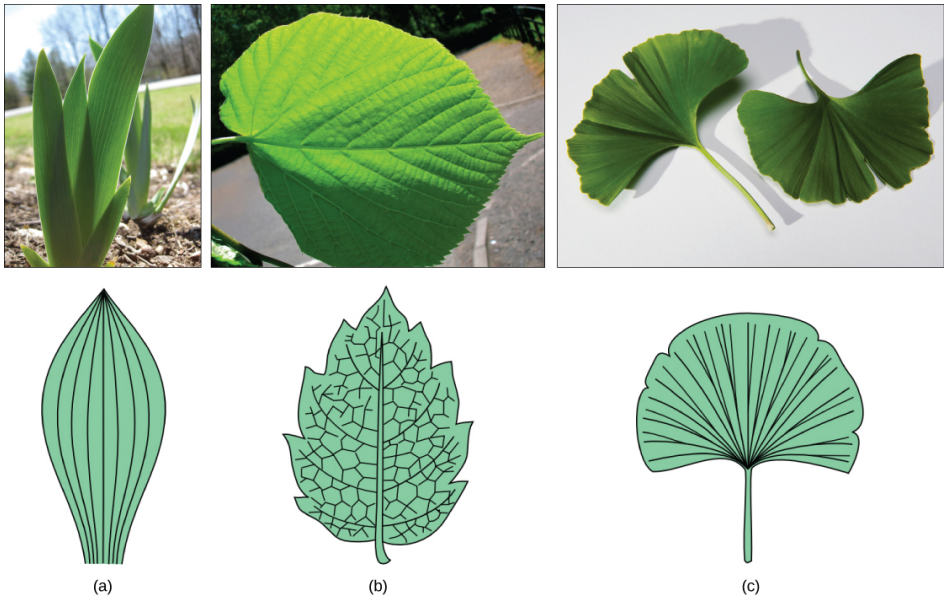


Figure 23.9 (a) Tulip (*Tulipa*), a monocot, has leaves with parallel venation. The netlike venation in this (b) linden (*Tilia cordata*) leaf distinguishes it as a eudicot. The (c) *Ginkgo biloba* tree has dichotomous venation. (credit a photo: modification of work by “Drewboy64”/Wikimedia Commons; credit b photo: modification of work by Roger Griffith; credit c photo: modification of work by "geishaboy500"/Flickr; credit abc illustrations: modification of work by Agnieszka Kwiecień)

Leaf Structure and Function

The outermost layer of the leaf is the epidermis; it is present on both sides of the leaf and is called the upper and lower epidermis, respectively. The epidermis helps in the regulation of gas exchange. It contains stomata (**Figure 23.10**): openings through which the exchange of gases takes place. Two guard cells surround each stoma, regulating its opening and closing.

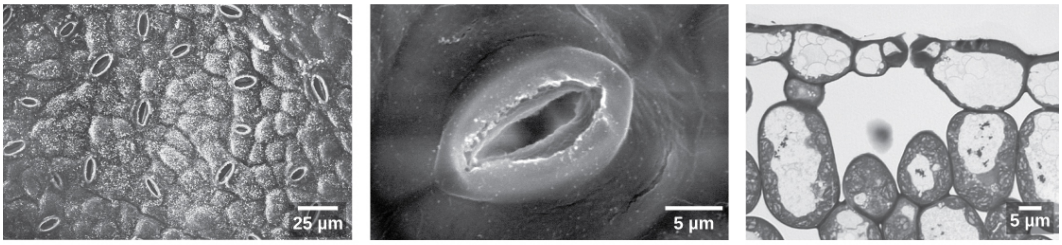


Figure 23.10 Visualized at 500x with a scanning electron microscope, several stomata are clearly visible on (a) the surface of this sumac (*Rhus glabra*) leaf. At 5,000x magnification, the guard cells of (b) a single stoma from lyre-leaved sand cress (*Arabidopsis lyrata)* have the appearance of lips that surround the opening. In this (c) light micrograph cross-section of an *A. lyrata* leaf, the guard cell pair is visible along with the large, sub-stomatal air space in the leaf. (credit: modification of work by Robert R. Wise; part c scale-bar data from Matt Russell)

When a plant has sufficient amount of water and sunlight, guard cells accumulate potassium (K^+) ions, and water flows into the guard cells by osmosis. The movement of water into the guard cells causes the cells to swell and become turgid. As the cells become more turgid, the inner thick cell wall does not stretch as much as the thin outer cell wall, and this causes a space to form between the two guard cells opening the stoma. When the plant closes that stoma, K^+ ions leave the guard cells, and water follows by osmosis. The movement of water out of the guard cells causes the cells to shrink and become flaccid thereby closing the stoma. **Figure 23.11** is an illustration of how guard cells open and close the stoma.

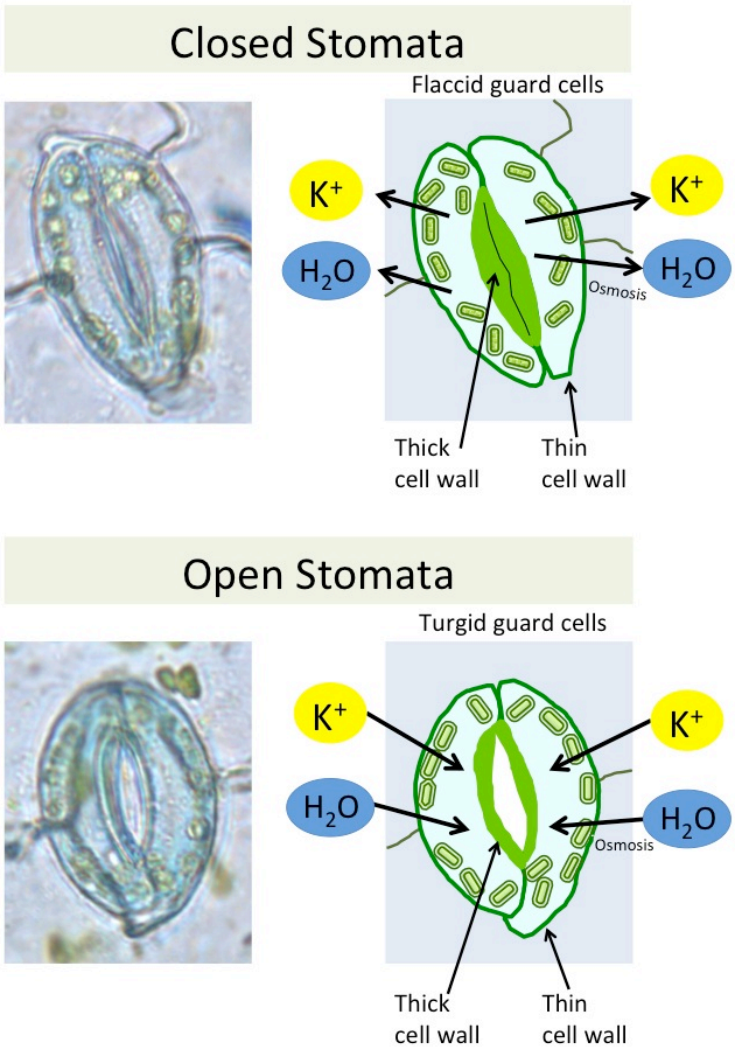
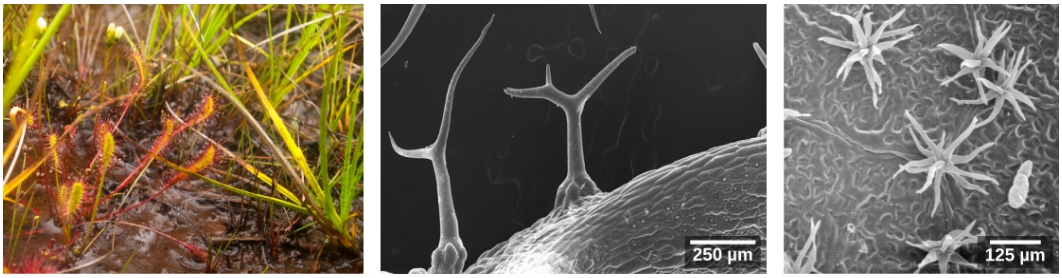


Figure 23.11 The movement of potassium ions (K^+) into and out of the guard cells allows the plant to either open or close its stomata. Work by Robert A. Bear

The movement of K^+ ions into or out of the guard cell as currently hypothesized is a passive process and related to the active transport of H^+ ions in the opposite direction of the movement of the K^+ ions. So, when H^+ ions are actively pumped out of the guard cell, K^+ ions move into the guard cell passively flowing down an electrical gradient created by the H^+ ions. Since the plant uses energy to open and close the stomata, the benefits of regulating the opening and closing of the stomata are greater than the energy expenditure of moving ions into and out of the guard cells. Plants actively regulate the movement of these ions and can respond rapidly to changes in the amount of sunlight, relative humidity and carbon dioxide.

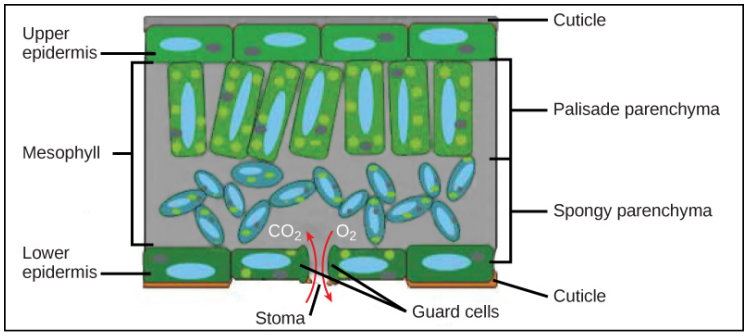
The epidermis is usually one cell layer thick; however, in plants that grow in very hot or very cold conditions, the epidermis may be several layers thick to protect against excessive water loss from transpiration. A waxy layer known as the **cuticle** covers the leaves of all plant species. The cuticle reduces the rate of water loss from the leaf surface. Other leaves may have small hairs (trichomes) on the leaf surface. Trichomes help to deter herbivory by restricting insect movements, or by storing toxic or bad-tasting compounds; they can also reduce the rate of transpiration by blocking air flow across the leaf surface (**Figure 23.12**).



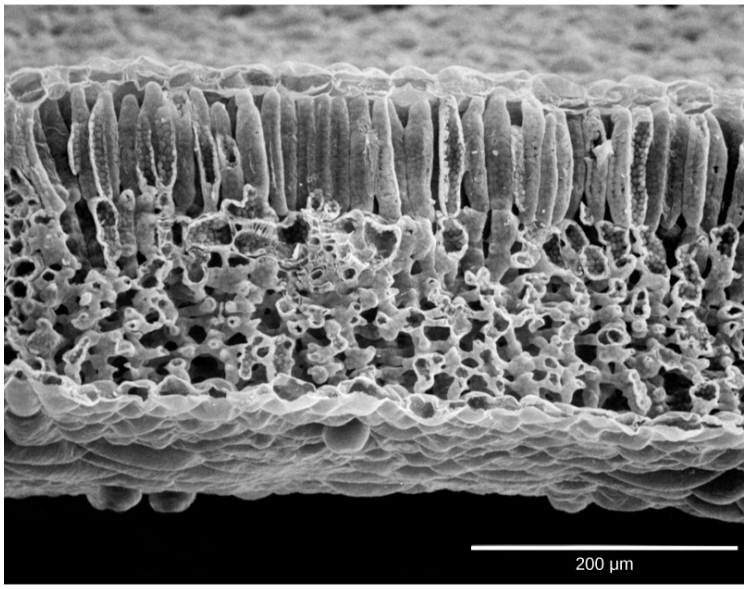
(a) (b) (c)

Figure 23.12 Trichomes give leaves a fuzzy appearance as in this (a) sundew (*Drosera* sp.). Leaf trichomes include (b) branched trichomes on the leaf of *Arabidopsis lyrata* and (c) multibranched trichomes on a mature *Quercus marilandica* leaf. (credit a: John Freeland; credit b, c: modification of work by Robert R. Wise; scale-bar data from Matt Russell)

Below the epidermis of eudicot leaves are layers of cells known as the mesophyll, or “middle leaf.” The mesophyll of most leaves typically contains two arrangements of parenchyma cells: the palisade parenchyma and spongy parenchyma (**Figure 23.13**). The palisade parenchyma (also called the palisade mesophyll) has column-shaped, tightly packed cells, and may be present in one, two, or three layers. Below the palisade parenchyma are loosely arranged cells of an irregular shape. These are the cells of the spongy parenchyma (or spongy mesophyll). The air space found between the spongy parenchyma cells allows gaseous exchange between the leaf and the outside atmosphere through the stomata. In aquatic plants, the intercellular spaces in the spongy parenchyma help the leaf float. Both layers of the mesophyll contain many chloroplasts. Guard cells are the only epidermal cells to contain chloroplasts.



(a)



(b)

Figure 23.13 In the (a) leaf drawing, the central mesophyll is sandwiched between an upper and lower epidermis. The mesophyll has two layers: an upper palisade layer comprised of tightly packed, columnar cells, and a lower spongy layer, comprised of loosely packed, irregularly shaped cells. Stomata on the leaf underside allow gas exchange. A waxy cuticle covers all aerial surfaces of land plants to minimize water loss. These leaf layers are clearly visible in the (b) scanning electron micrograph. The numerous small bumps in the palisade parenchyma cells are chloroplasts. Chloroplasts are also present in the spongy parenchyma, but are not as obvious. The bumps protruding from the lower surface of the leave are glandular trichomes, which differ in structure from the stalked trichomes in **Figure 23.12**. (credit b: modification of work by Robert R. Wise)

Like the stem, the leaf contains vascular bundles composed of xylem and phloem (**Figure 23.14**). The xylem consists of tracheids and vessels, which transport water and minerals to the leaves. The phloem transports the photosynthetic products from the leaf to the other parts of the plant. A single vascular bundle, no matter how large or small, always contains both xylem and phloem tissues.

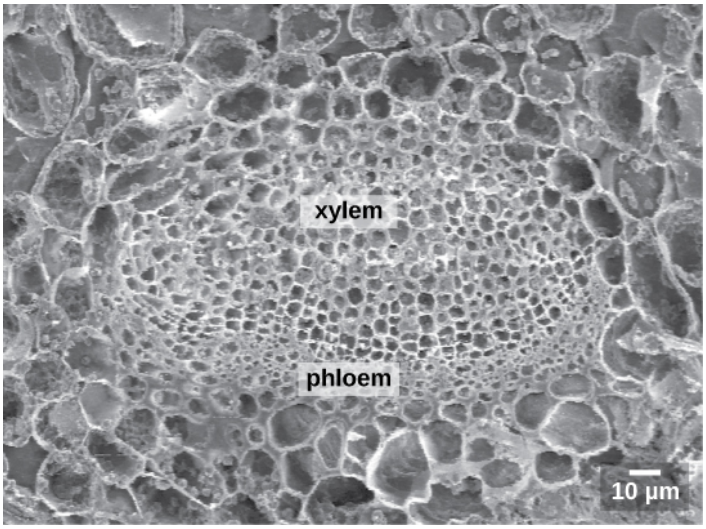


Figure 23.14 This scanning electron micrograph shows xylem and phloem in the leaf vascular bundle from the lyre-leaved sand cress (*Arabidopsis lyrata)*. (credit: modification of work by Robert R. Wise; scale-bar data from Matt Russell)

Leaf Adaptations

Coniferous plant species that thrive in cold environments, like spruce, fir, and pine, have leaves that are reduced in size and needle-like in appearance. These needle-like leaves have sunken stomata and a smaller surface area: two attributes that aid in reducing water loss. In hot climates, plants such as cacti have succulent leaves that help to conserve water. Many aquatic plants have leaves with wide lamina that can float on the surface of the water, and a thick waxy cuticle on the leaf surface that repels water.

evolution CONNECTION

Plant Adaptations in Resource-Deficient Environments

Roots, stems, and leaves are structured to ensure that a plant can obtain the required resources of sunlight, water, soil nutrients, carbon dioxide and oxygen. Some remarkable adaptations have evolved to enable plant species to thrive in less than ideal habitats, where one or more of these resources is in short supply.

In tropical rainforests, light is often scarce, since many trees and plants grow close together and block much of the sunlight from reaching the forest floor. Many tropical plant species have exceptionally broad leaves to maximize the capture of sunlight. Other species are epiphytes: plants that grow on other plants that serve as a physical support. Such plants are able to grow high up in the canopy atop the branches of other trees, where sunlight is more plentiful. Epiphytes live on rain and minerals collected in the branches and leaves of the supporting plant. Bromeliads (members of the pineapple family), ferns, and orchids are examples of tropical epiphytes (**Figure 23.15**). Many epiphytes have specialized tissues that enable them to efficiently capture and store water.

Figure 23.15 One of the most well known bromeliads is Spanish moss (*Tillandsia usneoides*), seen here in an oak tree. (credit: Kristine Paulus)

Some plants have special adaptations that help them to survive in nutrient-poor environments. Carnivorous plants, such as the Venus flytrap and the pitcher plant (**Figure 23.16**), grow in bogs where the soil is low in nitrogen. In these plants, leaves are modified to capture insects. The insect-capturing leaves may have evolved to provide these plants with a supplementary source of much-needed nitrogen.

(a) (b)

Figure 23.16 The (a) Venus flytrap has modified leaves that can capture insects. When an unlucky insect touches the trigger hairs inside the leaf, the trap suddenly closes. The opening of the (b) pitcher plant is lined with a slippery wax. Insects crawling on the lip slip and fall into a pool of water in the bottom of the pitcher, where they are digested by bacteria. The plant then absorbs the smaller molecules. (credit a: modification of work by Peter Shanks; credit b: modification of work by Tim Mansfield)

Many swamp plants have adaptations that enable them to thrive in wet areas, where their roots grow submerged underwater. In these aquatic areas, the soil is unstable and little oxygen is available to reach the roots. Trees such as mangroves (*Rhizophora* sp.) growing in coastal waters produce aboveground roots that help support the tree (**Figure 23.17**). Some species of mangroves, as well as cypress trees, have pneumatophores: upward-growing roots containing pores and pockets of tissue specialized for gas exchange. Wild rice is an aquatic plant with large air spaces in the root cortex. The air-filled tissue—called aerenchyma—provides a path for oxygen to diffuse down to the root tips, which are embedded in oxygen-poor bottom sediments.

(a) (b) (c)

Figure 23.17 The branches of (a) mangrove trees develop aerial roots, which descend to the ground and help to anchor the trees. (b) Cypress trees and some mangrove species have upward-growing roots called pneumatophores that are involved in gas exchange. Aquatic plants such as (c) wild rice have large spaces in the root cortex called aerenchyma, visualized here using scanning electron microscopy. (credit a: modification of work by Roberto Verzo; credit b: modification of work by Duane Burdick; credit c: modification of work by Robert R. Wise)

23.2 | Transport of Water and Solutes in Plants

> “"What else is the function of a forest, first and foremost, if not a place to do this: to capture and filter water and merge with sunlight, to create intricate being, intricate matter?"”

Rick Bass, "The Larch", Orion Magazine, September/October 2012

"The ability of trees to suck water from roots to leaves, sometimes to heights of over a hundred meters, is remarkable given the absence of any mechanical pump" wrote Harvey R. Brown in a review article in *Physics in Perspective* in 2013. Scientists have discovered quite a lot about this remarkable process, but that makes it no less remarkable. The structure of plant roots, stems, and leaves facilitates the transport of water, nutrients, and photosynthates throughout the plant. The phloem and xylem are the main tissues responsible for this movement. Water potential, evapotranspiration, and stomatal regulation influence how water and nutrients are transported in plants. To understand how these processes work, we must first understand the energetics of water potential.

Water Potential

Plants are phenomenal hydraulic engineers. Using only the basic laws of physics and the simple manipulation of potential energy, plants can move water to the top of a 116-meter-tall tree (**Figure 23.18a**). Plants can also use hydraulics to generate enough force to split rocks and buckle sidewalks (**Figure 23.18b**). Plants achieve this because of water potential.

(a) (b)

Figure 23.18 With heights nearing 116 meters, (a) coast redwoods (*Sequoia sempervirens*) are the tallest trees in the world. Plant roots can easily generate enough force to (b) buckle and break concrete sidewalks, much to the dismay of homeowners and city maintenance departments. (credit a: modification of work by Bernt Rostad; credit b: modification of work by Pedestrians Educating Drivers on Safety, Inc.)

Water potential is a measure of the potential energy in water. Plant physiologists are not interested in the energy in any one particular aqueous system, but are very interested in water movement between two systems. In practical terms, therefore, water potential is the difference in potential energy between a given water sample and pure water (at atmospheric pressure and ambient temperature). Water potential is expressed in units of pressure (pressure is a form of energy) called megapascals (MPa). The potential of pure water is, by convenience of definition, designated a value of zero (even though pure water contains plenty of potential energy, that energy is ignored). Water potential values for the water in a plant root, stem, or leaf are therefore expressed relative to pure H_2O.

The water potential in plant solutions is influenced by solute concentration, pressure, gravity, and factors called matrix effects. “System” can refer to the water potential of the soil water, root water, stem water, leaf water or the water in the atmosphere: whichever aqueous system is under consideration. As the individual components change, they raise or lower the total water potential of a system. When this happens, water moves to equilibrate, moving from the system or compartment with a higher water potential to the system or compartment with a lower water potential. This brings the difference in water potential between the two systems back to zero. Therefore, for water to move through the plant from the soil to the air (a process called transpiration), the water potential of the soil water > root water > stem water > leaf water > atmosphere water.

Water only moves in response to changes in water potential, not in response to the individual components. However, because the individual components influence the total water potential of the system, by manipulating the individual components solute concentration, a plant can control water movement.

Solute Potential

Solute potential, also called osmotic potential, is negative in a plant cell and zero in distilled water. Typical values for cell cytoplasm are –0.5 to –1.0 MPa. Solutes reduce water potential (resulting in a negative water potential) by consuming some of the potential energy available in the water. The internal water potential of a plant cell is more negative than pure water because of the cytoplasm’s high solute content (**Figure 23.19**). Because of this difference in water potential water will move from the soil into a plant’s root cells via the process of osmosis. This is why solute potential is sometimes called osmotic

potential.

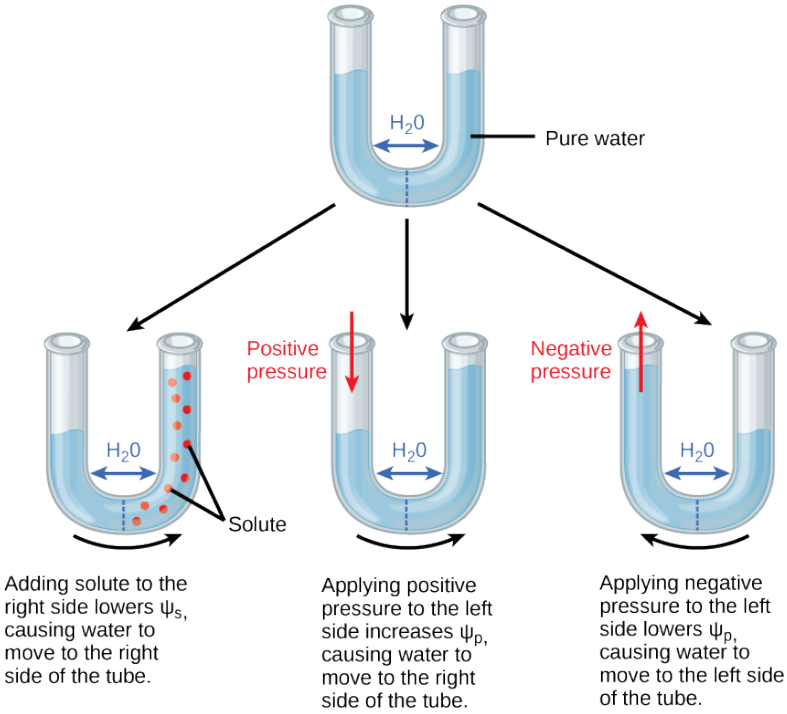


Figure 23.19 In this example with a semipermeable membrane between two aqueous systems, water will move from a region of higher to lower water potential until equilibrium is reached. Solutes (Ψ_s), pressure (Ψ_p), and gravity (Ψ_g) influence total water potential for each side of the tube ($\Psi_{total}^{right\ or\ left}$), and therefore, the difference between Ψ_{total} on each side ($\Delta\Psi$). (Ψ_m , the potential due to interaction of water with solid substrates, is ignored in this example because glass is not especially hydrophilic). Water moves in response to the difference in water potential between two systems (the left and right sides of the tube).

Pressure Potential

Pressure potential, also called turgor potential, may be positive or negative (**Figure 23.19**). Because pressure is an expression of energy, the higher the pressure, the more potential energy in a system, and vice versa. An example of the effect of turgor pressure is the wilting of leaves and their restoration after the plant has been watered (**Figure 23.20**). Water lost from the leaves via transpiration is restored by uptake via the roots.

A plant can manipulate turgor pressure via its ability to manipulate solute concentration and by the process of osmosis. If a plant cell increases the cytoplasmic solute concentration water will move into the cell by osmosis. Turgor pressure is also under indirect plant control via the opening and closing of stomata. Stomatal openings allow water to evaporate from the leaf, reducing the water potential of the leaf and increasing it between the water in the leaf and the petiole, thereby allowing water to flow from the petiole into the leaf.

(a) (b)

Figure 23.20 When (a) total water potential (Ψ_{total}) is lower outside the cells than inside, water moves out of the cells and the plant wilts. When (b) the total water potential is higher outside the plant cells than inside, water moves into the cells, resulting in turgor pressure (Ψ_p) and keeping the plant erect. (credit: modification of work by Victor M. Vicente Selvas)

Gravity Potential

Gravity potential is always negative to zero in a plant with no height. It always removes or consumes potential energy from the system. The force of gravity pulls water downwards to the soil, reducing the total amount of potential energy in the water in the plan. The taller the plant, the taller the water column, and the more influence gravity has. On a cellular scale and in short plants, this effect is negligible and easily ignored. However, over the height of a tall tree like a coast redwood, the gravitational pull of –0.1 MPa is equivalent to an extra 1 MPa of resistance that must be overcome for water to reach the leaves of the tallest trees.

Matric Potential

The binding of water to a matrix of surrounding substances always removes or consumes potential energy from the system. Matric potential is similar to solute potential because it involves tying up the energy in an aqueous system by forming hydrogen bonds between the water and some other component. Matric potential is always negative to zero. In a dry system, it can be as low as –2 MPa in a dry seed, and it is zero in a water-saturated system.

Movement of Water and Minerals in the Xylem

Solutes, pressure, gravity, and matric potential are all important for the transport of water in plants. Water moves from an area of higher total water potential (higher Gibbs free energy) to an area of lower total water potential. Gibbs free energy is the energy associated with a chemical reaction that can be used to do work.

Transpiration is the loss of water from the plant through evaporation at the leaf surface. It is the main driver of water movement in the xylem. Transpiration is caused by the evaporation of water at the leaf–atmosphere interface; it creates negative pressure (tension) equivalent to –2 MPa at the leaf surface. This value varies greatly depending on the vapor pressure deficit, which can be negligible at high relative humidity and substantial at low relative humidity. Water from the roots is pulled up by this tension. At night, when stomata shut and transpiration stops, the water is held in the stem and leaf by the adhesion of water to the cell walls of the xylem vessels and tracheids, and the cohesion of water molecules to each other. This is called the **cohesion–tension theory** of sap ascent.

Inside the leaf at the cellular level, water on the surface of mesophyll cells saturates the cellulose microfibrils of the primary cell wall. The leaf contains many large intercellular air spaces for the exchange of oxygen for carbon dioxide, which is required for photosynthesis. The wet cell wall is exposed to this leaf internal air space, and the water on the surface of the cells evaporates into the air spaces, decreasing the thin film on the surface of the mesophyll cells. This decrease creates a greater tension on the water in the mesophyll cells (**Figure 23.21**), thereby increasing the pull on the water in the xylem vessels. The xylem vessels and tracheids are structurally adapted to cope with large changes in pressure. Rings in the vessels maintain their tubular shape, much like the rings on a vacuum cleaner hose keep the hose open while it is under pressure. Small perforations between vessel elements reduce the number and size of gas bubbles that can form via a process called cavitation. The formation of gas bubbles in xylem interrupts the continuous stream of water from the base to the top of the plant, causing a break termed an embolism in the flow of xylem sap. The taller the tree, the greater the tension forces needed to pull water, and the more cavitation events. In larger trees, the resulting embolisms can plug xylem vessels, making them non-functional.

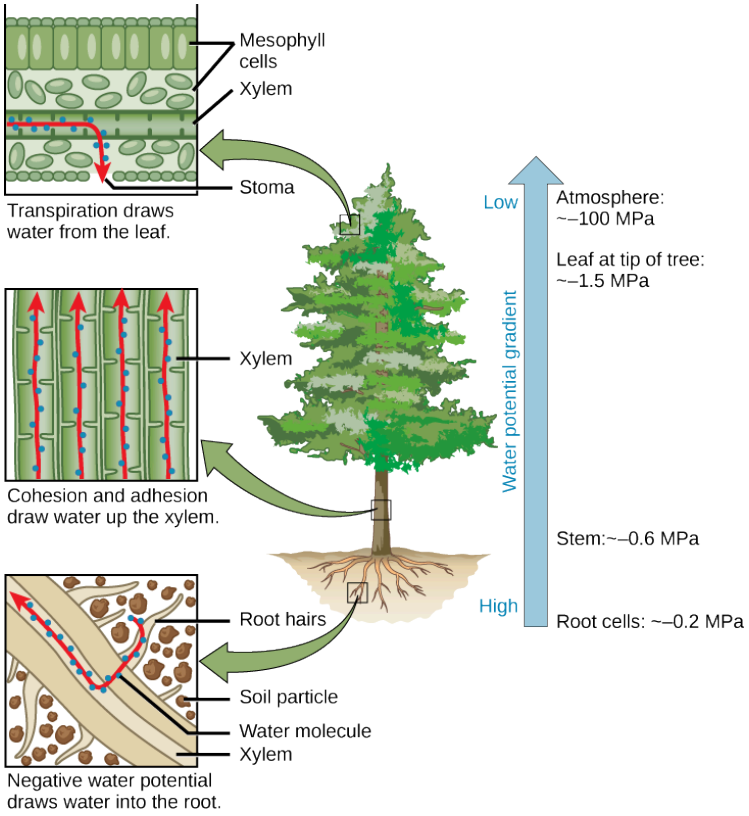


Figure 23.21 The cohesion–tension theory of sap ascent is shown. Evaporation from the mesophyll cells produces a negative water potential gradient that causes water to move upwards from the roots through the xylem.

Transpiration is a passive process, meaning that metabolic energy in the form of ATP is not required for water movement. The energy driving transpiration is the difference in energy between the water in the soil and the water in the atmosphere. However, transpiration is tightly controlled.

Control of Transpiration

The atmosphere to which the leaf is exposed drives transpiration, but also causes massive water loss from the plant. Up to 90 percent of the water taken up by roots may be lost through transpiration.

Leaves are covered by a waxy **cuticle** on the outer surface that prevents the loss of water. Regulation of transpiration, therefore, is achieved primarily through the opening and closing of stomata on the leaf surface. Stomata are surrounded by two specialized cells called guard cells, which open and close in response to environmental cues such as light intensity and quality, leaf water status, and carbon dioxide concentrations. Stomata must open to allow air containing carbon dioxide and oxygen to diffuse into the leaf for photosynthesis and respiration. When stomata are open, however, water vapor is lost to the external environment, increasing the rate of transpiration. Therefore, plants must maintain a balance between efficient photosynthesis and water loss.

Plants have evolved over time to adapt to their local environment and reduce transpiration (**Figure 23.22**). Desert plant (xerophytes) and plants that grow on other plants (epiphytes) have limited access to water. Such plants usually have a much thicker waxy cuticle than those growing in more moderate, well-watered environments (mesophytes). Aquatic plants (hydrophytes) also have their own set of anatomical and morphological leaf adaptations.

(a) (b)

(c) (d)

Figure 23.22 Plants are suited to their local environment. (a) Xerophytes, like this prickly pear cactus (*Opuntia sp.*) and (b) epiphytes such as this tropical *Aeschynanthus perrottetii* have adapted to very limited water resources. The leaves of a prickly pear are modified into spines, which lowers the surface-to-volume ratio and reduces water loss. Photosynthesis takes place in the stem, which also stores water. (b) *A. perottetii* leaves have a waxy cuticle that prevents water loss. (c) Goldenrod (*Solidago sp.*) is a mesophyte, well suited for moderate environments. (d) Hydrophytes, like this fragrant water lily (*Nymphaea odorata*), are adapted to thrive in aquatic environments. (credit a: modification of work by Jon Sullivan; credit b: modification of work by L. Shyamal/Wikimedia Commons; credit c: modification of work by Huw Williams; credit d: modification of work by Jason Hollinger)

Xerophytes and epiphytes often have a thick covering of trichomes or of stomata that are sunken below the leaf's surface. Trichomes are specialized hair-like epidermal cells that secrete oils and substances. These adaptations impede air flow across the stomatal pore and reduce transpiration. Multiple epidermal layers are also commonly found in these types of plants.

Transportation of Photosynthates in the Phloem

Plants need an energy source to grow. In seeds and bulbs, food is stored in polymers (such as starch) that are converted by metabolic processes into sucrose for newly developing plants. Once green shoots and leaves are growing, plants are able to produce their own food by photosynthesizing. The products of photosynthesis are called photosynthates, which are usually in the form of simple sugars such as sucrose.

Structures that produce photosynthates for the growing plant are referred to as **sources**. Sugars produced in sources, such as leaves, need to be delivered to growing parts of the plant via the phloem in a process called translocation. The points of sugar delivery, such as roots, young shoots, and developing seeds, are called **sinks**. Seeds, tubers, and bulbs can be either a source or a sink, depending on the plant's stage of development and the season.

The products from the source are usually moved (translocated) to the nearest sink through the phloem. For example, the highest leaves will send photosynthates upward to the growing shoot tip, whereas lower leaves will direct photosynthates downward to the roots. Intermediate leaves will send products in both directions, unlike the flow in the xylem, which is always unidirectional (soil to leaf to atmosphere). The pattern of photosynthate flow changes as the plant grows and develops. Photosynthates are directed primarily to the roots early on, to shoots and leaves during vegetative growth, and to seeds and fruits during reproductive development. They are also directed to tubers for storage.

Pressure Flow Model: Transport from Source to Sink

Photosynthates, such as sucrose, are produced in the mesophyll cells of photosynthesizing leaves. From there they are translocated through the phloem to where they are used or stored. Mesophyll cells are connected by cytoplasmic channels called plasmodesmata. Photosynthates move through these channels to reach phloem sieve-tube elements (STEs) in the vascular bundles. From the mesophyll cells, the photosynthates are loaded into the phloem STEs. The sucrose is actively transported against its concentration gradient (a process requiring ATP) into the phloem cells using the electrochemical potential of the proton gradient. This is coupled to the uptake of sucrose with a carrier protein called the sucrose-H^+ symporter.

Phloem STEs have reduced cytoplasmic contents, and are connected by a sieve plate with pores that allow for pressure-driven bulk flow, or translocation, of phloem sap. Companion cells are associated with STEs. They assist with metabolic activities and produce energy for the STEs (**Figure 23.23**).

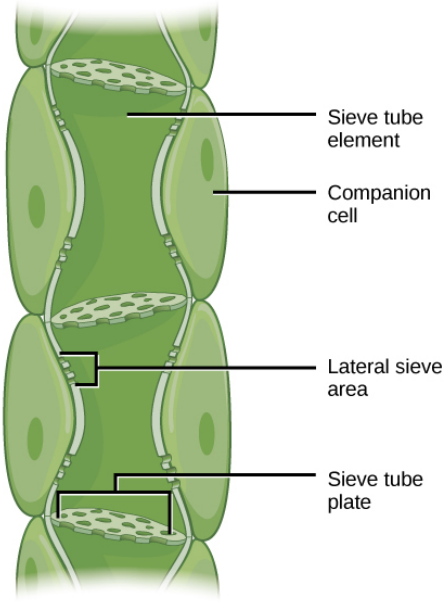


Figure 23.23 Phloem is comprised of cells called sieve-tube elements. Phloem sap travels through perforations called sieve tube plates. Neighboring companion cells carry out metabolic functions for the sieve-tube elements and provide them with energy. Lateral sieve areas connect the sieve-tube elements to the companion cells.

Once in the phloem, the photosynthates are translocated to the closest sink. Phloem sap is an aqueous solution that contains up to 30 percent sugar, minerals, amino acids, and plant growth regulators. The high percentage of sugar causes water to move by osmosis from the adjacent xylem into the phloem tubes, thereby increasing pressure. This increase in total water potential causes the bulk flow of phloem from source to sink (**Figure 23.24**). Sucrose concentration in the sink cells is lower than in the phloem STEs because the sink sucrose has been metabolized for growth, or converted to starch for storage or other polymers, such as cellulose, for structural integrity. Unloading at the sink end of the phloem tube occurs by either diffusion or active transport of sucrose molecules from an area of high concentration to one of low concentration. Water diffuses from the phloem by osmosis and is then transpired or recycled via the xylem back into the phloem sap. This concept of how photosynthates move from the source to the sink is called the **pressure flow model**.

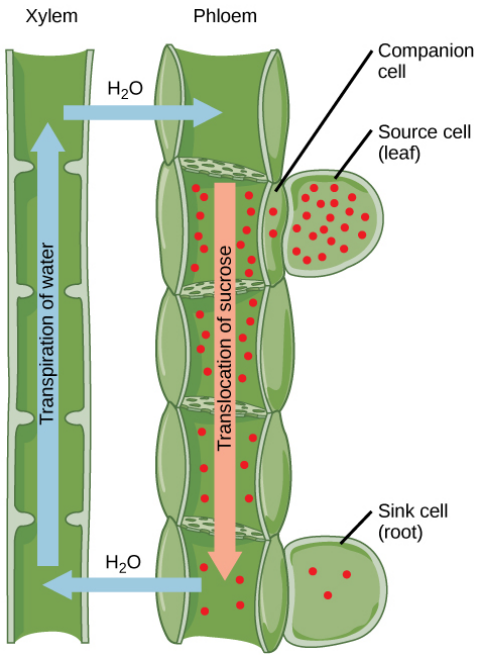


Figure 23.24 Sucrose is actively transported from source cells into companion cells and then into the sieve-tube elements. This reduces the water potential, which causes water to enter the phloem from the xylem. The resulting positive pressure forces the sucrose-water mixture down toward the roots, where sucrose is unloaded. Transpiration causes water to return to the leaves through the xylem vessels.

23.3 | Nutritional Requirements of Plants

> "I died as mineral and became a plant,"
>
> "I died as plant and rose to animal,"
>
> "I died as animal and became man."
>
> "Why should I fear? When was I less by dying?"

Jalal ad-Din ar-Rumi, Sufi poet (1207-73)

Rumi's musings on minerals and plants contain a kernel of truth. Minerals become incorporated into the plant body, and are critically important in plant function. Plants are unique organisms that can absorb nutrients and water through their root system, as well as carbon dioxide from the atmosphere. Soil quality and climate are the major determinants of plant distribution and growth. The combination of soil nutrients, water, and carbon dioxide, along with sunlight, allows plants to grow.

The Chemical Composition of Plants

Since plants require nutrients in the form of elements such as carbon and potassium, it is important to understand the chemical composition of plants. The majority of volume in a plant cell is water; it typically comprises 80 to 90 percent of the plant's total weight. Soil is the water source for land plants, and can be an abundant source of water, even if it appears dry. Plant roots absorb water from the soil through root hairs and transport it up to the leaves through the xylem. As water vapor is lost from the leaves, the process of transpiration and the polarity of water molecules (which enables them to form hydrogen bonds) draws more water from the roots up through the plant to the leaves (**Figure 23.25**). Plants need water to

support cell structure, for metabolic functions, to carry nutrients, and for photosynthesis.

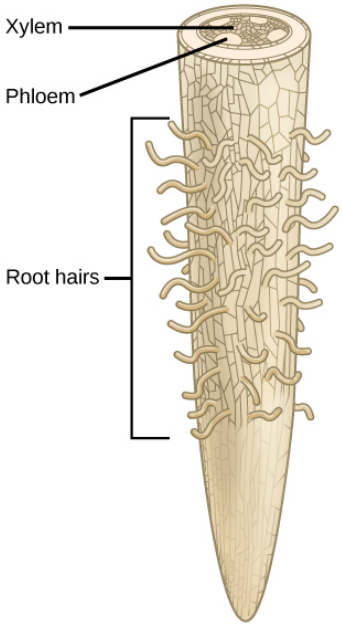


Figure 23.25 Water is absorbed through the root hairs and moves up the xylem to the leaves.

Plant cells need essential substances, collectively called nutrients, to sustain life. Plant nutrients may be composed of either organic or inorganic compounds. An **organic compound** is a chemical compound that contains carbon, such as carbon dioxide obtained from the atmosphere. Carbon that was obtained from atmospheric CO_2 composes the majority of the dry mass within most plants. An **inorganic compound** does not contain carbon and is not part of, or produced by, a living organism. Inorganic substances, which form the majority of the soil solution, are commonly called minerals: those required by plants include nitrogen (N) and potassium (K) for structure and regulation.

Essential Nutrients

Plants require only light, water and about 20 elements to support all their biochemical needs: these 20 elements are called essential nutrients (Table 23.1). For an element to be regarded as **essential**, three criteria are required: 1) a plant cannot complete its life cycle without the element; 2) no other element can perform the function of the element; and 3) the element is directly involved in plant nutrition.

Essential Elements for Plant Growth

Macronutrients	Micronutrients
Carbon (C)	Iron (Fe)
Hydrogen (H)	Manganese (Mn)
Oxygen (O)	Boron (B)
Nitrogen (N)	Molybdenum (Mo)
Phosphorus (P)	Copper (Cu)
Potassium (K)	Zinc (Zn)
Calcium (Ca)	Chlorine (Cl)
Magnesium (Mg)	Nickel (Ni)
Sulfur (S)	Cobalt (Co)
	Sodium (S)

Table 23.1

Essential Elements for Plant Growth

Macronutrients	Micronutrients
	Silicon (Si)

Table 23.1

Macronutrients and Micronutrients

The essential elements can be divided into two groups: macronutrients and micronutrients. Nutrients that plants require in larger amounts are called **macronutrients**. About half of the essential elements are considered macronutrients: carbon, hydrogen, oxygen, nitrogen, phosphorus, potassium, calcium, magnesium and sulfur. The first of these macronutrients, carbon (C), is required to form carbohydrates, proteins, nucleic acids, and many other compounds; it is therefore present in all macromolecules. On average, the dry weight (excluding water) of a cell is 50 percent carbon. As shown in **Figure 23.26**, carbon is a key part of plant biomolecules.

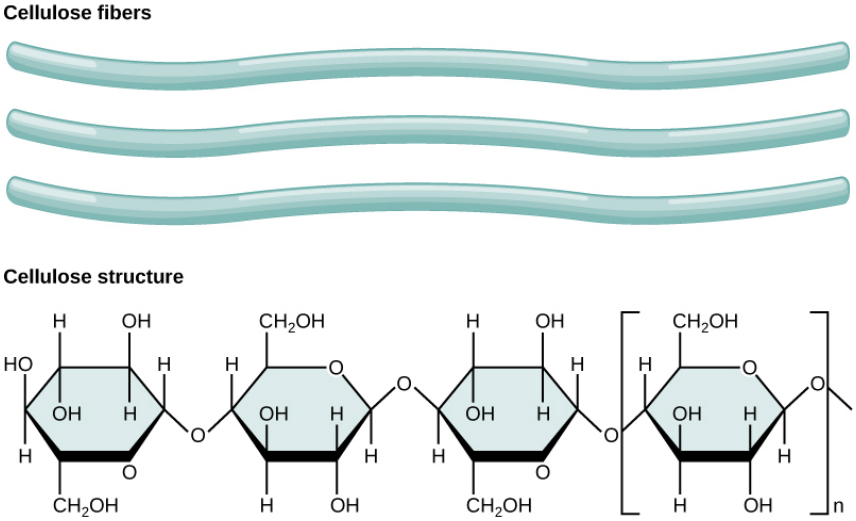


Figure 23.26 Cellulose, the main structural component of the plant cell wall, makes up over thirty percent of plant matter. It is the most abundant organic compound on earth. Plants are able to make their own cellulose, but need carbon from the soil to do so.

The next most abundant element in plant cells is nitrogen (N); it is part of proteins and nucleic acids. Nitrogen is also used in the synthesis of some vitamins. Hydrogen and oxygen are macronutrients that are part of many organic compounds, and also form water. Oxygen is necessary for cellular respiration; plants use oxygen to store energy in the form of ATP. Phosphorus (P), another macromolecule, is necessary to synthesize nucleic acids and phospholipids. As part of ATP, phosphorus enables food energy to be converted into chemical energy through oxidative phosphorylation. Likewise, light energy is converted into chemical energy during photophosphorylation in photosynthesis, and into chemical energy to be extracted during respiration. Sulfur is part of certain amino acids, such as cysteine and methionine, and is present in several coenzymes. Sulfur also plays a role in photosynthesis as part of the electron transport chain, where hydrogen gradients play a key role in the conversion of light energy into ATP. Potassium (K) is important because of its role in regulating stomatal opening and closing. As the openings for gas exchange, stomata help maintain a healthy water balance; a potassium ion pump supports this process.

Magnesium (Mg) and calcium (Ca) are also important macronutrients. The role of calcium is twofold: to regulate nutrient transport, and to support many enzyme functions. Magnesium is important to the photosynthetic process. These minerals, along with the micronutrients, which are described below, also contribute to the plant's ionic balance.

In addition to macronutrients, organisms require various elements in small amounts. These **micronutrients**, or trace elements, are present in very small quantities. They include boron (B), chlorine (Cl), manganese (Mn), iron (Fe), zinc (Zn), copper (Cu), molybdenum (Mo), nickel (Ni), silicon (Si), and sodium (Na).

Deficiencies in any of these nutrients—particularly the macronutrients—can adversely affect plant growth (**Figure 23.27**. Depending on the specific nutrient, a lack can cause stunted growth, slow growth, or chlorosis (yellowing of the leaves). Extreme deficiencies may result in leaves showing signs of cell death.

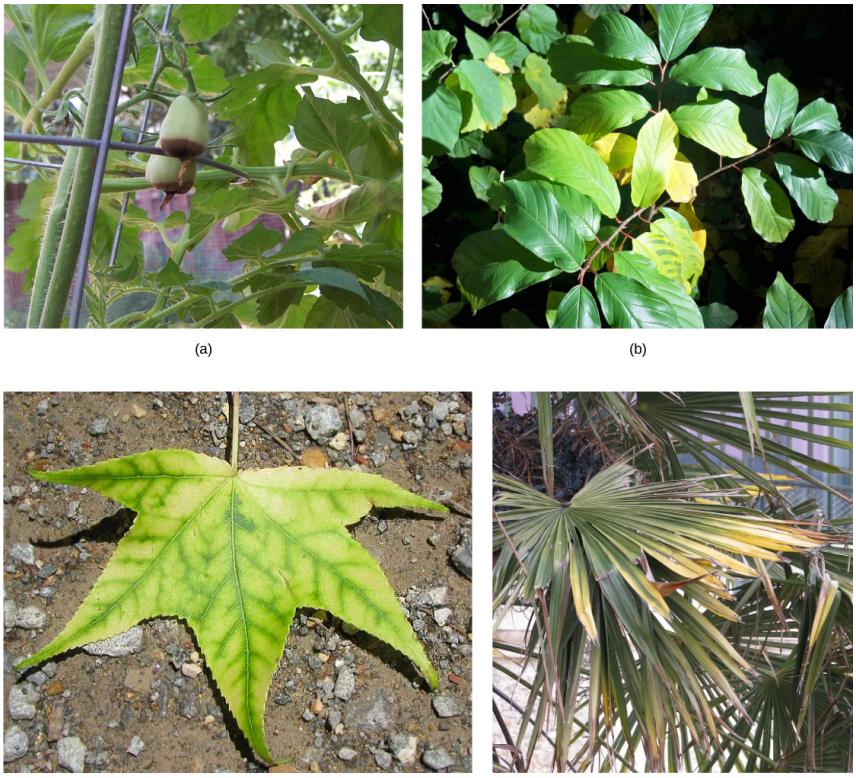


Figure 23.27 Nutrient deficiency is evident in the symptoms these plants show. This (a) grape tomato suffers from blossom end rot caused by calcium deficiency. The yellowing in this (b) *Frangula alnus* results from magnesium deficiency. Inadequate magnesium also leads to (c) intervenal chlorosis, seen here in a sweetgum leaf. This (d) palm is affected by potassium deficiency. (credit c: modification of work by Jim Conrad; credit d: modification of work by Malcolm Manners)

23.4 | Nutritional Adaptations of Plants

B

Introduction

> “The fixation of atmospheric nitrogen is one of the great discoveries, awaiting the genius of chemists.”

Sir William Crookes Presidential Address to the British Association for the Advancement of Science, 1898

Of course, plants discovered how to fix nitrogen (convert N_2 gas to a form that can be incorporated in biological molecules such as nucleic acids and proteins) a long time ago. Or rather, they recruited nitrogen-fixing bacteria into a mutualistic relationship, where the bacteria fix nitrogen and provide nitrogen-containing compounds to the plant in exchange for food and shelter.

Plants may also enlist the help of other microbial partners in nutrient acquisition. Particular species of bacteria and fungi have evolved along with certain plants to create a mutualistic symbiotic relationship with roots. This improves the nutrition of both the plant and the microbe. The formation of bacteria-containing nodules on the roots for nitrogen fixation, or the association of mycorrhizal fungi with roots for enhanced nutrient and water uptake, can be considered among the nutritional adaptations of plants. However, these are not the only type of adaptations that we may find; many plants have other adaptations that allow them to thrive under specific conditions.

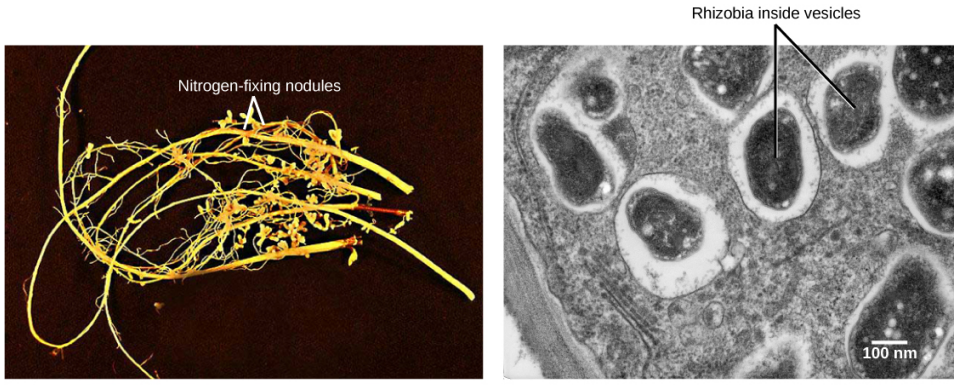
Nitrogen Fixation: Root and Bacteria Interactions

Nitrogen is an important macronutrient because it is part of nucleic acids and proteins. Atmospheric nitrogen, which is the diatomic molecule N_2, or dinitrogen, is the largest pool of nitrogen in terrestrial ecosystems. However, plants cannot take advantage of this nitrogen because they do not have the necessary enzymes to convert it into biologically useful forms. However, nitrogen can be "fixed," which means that it can be converted to ammonia (NH_3) through biological, physical, or chemical processes. As you have learned, biological nitrogen fixation is the conversion of atmospheric nitrogen (N_2) into ammonia (NH_3), exclusively carried out by prokaryotes such as soil bacteria or cyanobacteria. Biological processes contribute 65 percent of the nitrogen used in agriculture.

(a) (b) (c)

Figure 23.28 Some common edible legumes—like (a) peanuts, (b) beans, and (c) chickpeas—are able to interact symbiotically with soil bacteria that fix nitrogen. (credit a: modification of work by Jules Clancy; credit b: modification of work by USDA)

Nitrogen-fixing soil bacteria, collectively called **rhizobia**, symbiotically interact with legume roots to form specialized structures called **nodules**, in which nitrogen fixation takes place. This process entails the reduction of atmospheric nitrogen to ammonia, by means of the enzyme nitrogenase. Therefore, using rhizobia is a natural and environmentally friendly way to fertilize plants, as opposed to chemical fertilization that uses a nonrenewable resource, such as natural gas. Through symbiotic nitrogen fixation, the plant benefits from using an endless source of nitrogen from the atmosphere. The process simultaneously contributes to soil fertility because the plant root system leaves behind some of the biologically available nitrogen. As in any symbiosis, both organisms benefit from the interaction: the plant obtains ammonia, and bacteria obtain carbon compounds generated through photosynthesis, as well as a protected niche in which to grow (**Figure 23.29**).



(a) (b)

Figure 23.29 Soybean roots contain (a) nitrogen-fixing nodules. Cells within the nodules are infected with *Bradyrhyzobium japonicum,* a rhizobia or "root-loving" bacterium. The bacteria are encased in (b) vesicles inside the cell, as can be seen in this transmission electron micrograph. (credit a: modification of work by USDA; credit b: modification of work by Louisa Howard, Dartmouth Electron Microscope Facility; scale-bar data from Matt Russell)

Mycorrhizae: The Symbiotic Relationship between Fungi and Roots

A nutrient depletion zone can develop when there is rapid soil solution uptake, low nutrient concentration, low diffusion rate, or low soil moisture. These conditions are very common; therefore, most plants rely on fungi to facilitate the uptake of minerals from the soil. Fungi form symbiotic associations called mycorrhizae with plant roots, in which the fungi actually are integrated into the physical structure of the root. The fungi colonize the living root tissue during active plant growth.

Through mycorrhization, the plant obtains mainly phosphate and other minerals, such as zinc and copper, from the soil. The fungus obtains nutrients, such as sugars, from the plant root (**Figure 23.30**). Mycorrhizae help increase the surface area of

the plant root system because hyphae, which are narrow, can spread beyond the nutrient depletion zone. Hyphae can grow into small soil pores that allow access to phosphorus that would otherwise be unavailable to the plant. The beneficial effect on the plant is best observed in poor soils. The benefit to fungi is that they can obtain up to 20 percent of the total carbon accessed by plants. Mycorrhizae functions as a physical barrier to pathogens. It also provides an induction of generalized host defense mechanisms, and sometimes involves production of antibiotic compounds by the fungi.

Figure 23.30 Root tips proliferate in the presence of mycorrhizal infection, which appears as off-white fuzz in this image. (credit: modification of work by Nilsson et al., BMC Bioinformatics 2005)

There are two types of mycorrhizae: ectomycorrhizae and endomycorrhizae. Ectomycorrhizae form an extensive dense sheath around the roots, called a mantle. Hyphae from the fungi extend from the mantle into the soil, which increases the surface area for water and mineral absorption. This type of mycorrhizae is found in forest trees, especially conifers, birches, and oaks. Endomycorrhizae, also called arbuscular mycorrhizae, do not form a dense sheath over the root. Instead, the fungal mycelium is embedded within the root tissue. Endomycorrhizae are found in the roots of more than 80 percent of terrestrial plants.

Nutrients from Other Sources

Some plants cannot produce their own food and must obtain their nutrition from outside sources. This may occur with plants that are parasitic or saprophytic. Some plants are mutualistic symbionts, epiphytes, or insectivorous.

Plant Parasites

A parasitic plant depends on its host for survival. Some parasitic plants have no leaves. An example of this is the dodder (**Figure 23.31**), which has a weak, cylindrical stem that coils around the host and forms suckers. From these suckers, cells invade the host stem and grow to connect with the vascular bundles of the host. The parasitic plant obtains water and nutrients through these connections. The plant is a total parasite (a holoparasite) because it is completely dependent on its host. Other parasitic plants (hemiparasites) are fully photosynthetic and only use the host for water and minerals. There are about 4,100 species of parasitic plants.

Figure 23.31 The dodder is a holoparasite that penetrates the host's vascular tissue and diverts nutrients for its own growth. Note that the vines of the dodder, which has white flowers, are beige. The dodder has no chlorophyll and cannot produce its own food. (credit: "Lalithamba"/Flickr)

Saprophytes

A **saprophyte** is a plant that does not have chlorophyll and gets its food from dead matter, similar to bacteria and fungi (note that fungi are often called saprophytes, which is incorrect, because fungi are not plants). Plants like these use enzymes to convert organic food materials into simpler forms from which they can absorb nutrients (**Figure 23.32**). Most saprophytes do not directly digest dead matter: instead, they parasitize fungi that digest dead matter, or are mycorrhizal, ultimately obtaining photosynthate from a fungus that derived photosynthate from its host. Saprophytic plants are uncommon; only a few species are described.

Figure 23.32 Saprophytes, like this Spotted Coralroot (*Corallorhiza maculata)*, obtain their food from the mycelium of soil fungi, and do not have chlorophyll. The Spotted Coralroot is found in montane forests in northern and western North America and Central America. This specimen was photographed in the Pecos Wilderness area in New Mexico. (credit: photo by D. A. Rintoul)

Epiphytes

An epiphyte is a plant that grows on other plants, but is not dependent upon the other plant for nutrition (**Figure 23.33**). Epiphytes have two types of roots: clinging aerial roots, which absorb nutrients from humus that accumulates in the crevices of trees; and aerial roots, which absorb moisture from the atmosphere.

Figure 23.33 These epiphyte plants grow in the main greenhouse of the *Jardin des Plantes* in Paris.

Insectivorous Plants

An insectivorous plant has specialized leaves to attract and digest insects. You are probably familiar with the Venus flytrap, which has leaves that work as snap-traps . Other kinds of carnivorous plants, such as the sundews, are decorated with glands that secrete a sticky fluid which both attracts and then later digests insects that become ensnared in the sticky fluid. And still another group includes the pitcher plants, which typically catch and hold rainwater in a pitcher-shaped organ. The plant secretes nutrients as well as digestive enzymes into the trapped water. Insects are attracted to, and then fall into, the pool of fluid; they cannot easily escape due to the downward-facing hairs that line the inside of the "pitcher". One unique kind of pitcher plant is the California Pitcher Plant (*Darlingtonia californica*, **Figure 23.34**), found in bogs and seeps in the mountains of northern California and southern Oregon. The plant, also known as the Cobra Lily, is unique in that it does not trap rainwater in the pitcher, but rather pumps it up from the roots. A tiny entrance hole underneath the "head" of the cobra-shaped structure allows insects in, but they cannot get back out once they are trapped in the digestive fluid stored below. All carnivorous plants are found in nitrogen-poor soils, and obtain the bulk of this critical nutrient from the bodies of insects and other animals that they trap and digest.

Figure 23.34 A California Pitcher Plant has specialized leaves to trap insects. This specimen was photographed in the Trinity Alps Wilderness in northern California. (credit: photo by D. A. Rintoul)

24 | PHOTOSYNTHESIS, GLOBAL CLIMATE CHANGE, AND FOOD PRODUCTION

24.1 | Photosynthetic Pathways

> “You can't have a light without a dark to stick it in”

." Arlo Guthrie, American musician

In a previous module, you learned about photosynthesis, the mechanism plants use to convert solar energy into chemical energy. The light energy captured is used to make ATP and NADPH, which is then used to reduce carbon from a simple form (CO_2) into a more complex form (sugars). The first step of the Calvin cycle is the fixation of carbon dioxide to RuBP, and the plants that only use this mechanism of carbon fixation are called C_3 plants. About 85% of the plant species on the planet are C_3 plants; some examples are rice, wheat, soybeans and all trees.

The process of photosynthesis has a theoretical efficiency of 30% (i.e., the maximum amount of chemical energy output would be only 30% of the solar energy input), but in reality the efficiency is much lower. It is only about 3% on cloudy days. Why is so much solar energy lost? There are a number of factors contributing to this energy loss, and one metabolic pathway that contributes to this low efficiency is photorespiration. During photorespiration, the key photosynthetic enzyme Rubisco (ribulose-1,5-bisphosphate carboxylase oxygenase) uses O_2 as a substrate instead of CO_2. This process uses up a considerable amount of energy without making sugars (**Figure 24.1**). When a plant has its stomata open (CO_2 is diffusing in while O_2 and water are diffusing out), photorespiration is minimized because Rubisco has a higher affinity for CO_2 than for O_2 when air temperatures are below 30°C (86°F). However, when a plant closes its stomata during times of water stress and O_2 from photosynthesis builds up inside the cell, the rate of photorespiration increases because O_2 is now more abundant inside the mesophyll. So, there is a tradeoff. Plants can leave the stomata open and risk drying out, or they can close the stomata, thereby reducing the uptake of CO_2, and decreasing the efficiency of photosynthesis. In addition, Rubisco has a higher affinity for O_2 when temperatures increase, which means that C_3 plants use more energy (ATP) for photorespiration at higher temperatures.

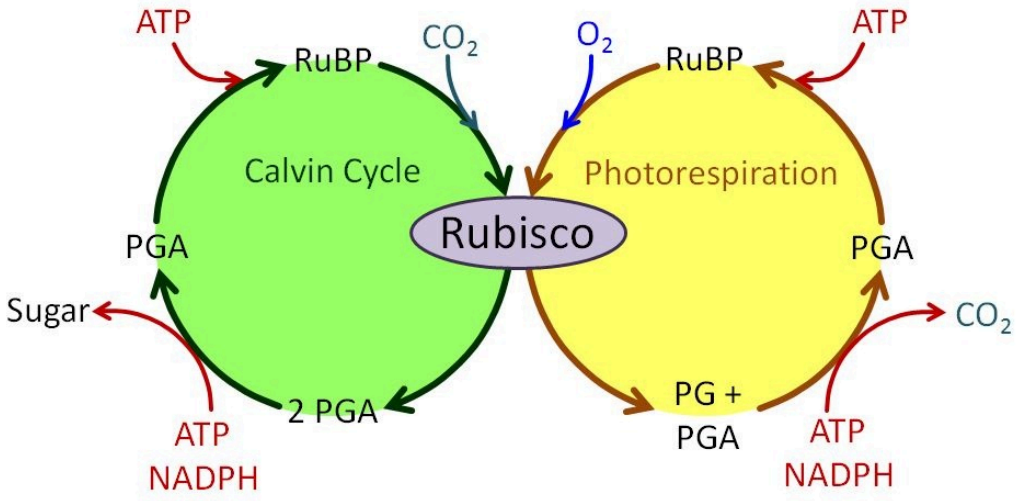


Figure 24.1 A comparison of photorespiration and carbon fixation in C3 plants. During photorespiration, O_2 is bound to RuBP and forms phosphoglycolate (PG) and Phosphoglycerate (PGA), PG then undergoes an number energy requiring reactions releasing CO2. Work by Eva Horne.

Evolutionarily speaking, why is photorespiration still around? One hypothesis is that it is evolutionary baggage from a time when the atmosphere had a lower O_2 concentration than it does today. In other words, when Rubisco first evolved millions of years ago, the O_2 concentration was so low that excluding O_2 from its binding site had little or no influence on the efficiency of photosynthesis. The modern Rubisco retains some of its ancestral affinity for O_2, which leads to the energy costs associated with photorespiration. However, plant cell physiologists are discovering that there might be some metabolic benefits associated with photorespiration, which would help explain why this seemingly wasteful pathway is still found in plants. Adding to the dilemma is the fact that when plant geneticists "knock out" Rubisco's ability to fix O_2, Rubisco also loses its ability to fix CO_2. It is possible that the active site of this enzyme cannot be engineered, by artificial or natural selection, so that it exclusively binds CO_2 and not O_2.

C_4 plant and CAM Pathways as a Means of Reducing Photorespiration

The C_4 and CAM pathways for fixing CO_2 are two adaptations that improve the efficiency of photosynthesis, by ensuring that Rubisco encounters high CO_2 concentrations and thus reduces photorespiration. These two photosynthetic adaptations for fixing CO_2 have evolved independently a number of times in species that evolved from wet and dry, but typically warm climates. Why have these mechanisms evolved independently so many times? Plants that minimize photorespiration may have a significant competitive advantage, because a considerable amount of energy (in the form of ATP) is lost in plants during photorespiration. In many environments, plants that use solar energy more efficiently should out-compete those which are less efficient.

C_4 Pathway

In C_4 plants, the light-dependent reactions and the Calvin cycle are physically separated, with the light-dependent reactions occurring in the mesophyll cells and the Calvin cycle occurring in special cells that surround the veins in the leaves. These cells are called bundle-sheath cells. How does this work? Atmospheric CO_2 is fixed in the mesophyll cells as a simple 4-carbon organic acid (malate) by an enzyme that has no affinity for O_2. Malate is then transported to the bundle-sheath cells. Inside the bundle sheath, malate is oxidized to a 3-C organic acid, and in the process, 1 molecule of CO_2 is produced from every malate molecule (**Figure 24.2**). The CO_2 is then fixed by Rubisco into sugars, via the Calvin cycle, exactly as in C_3 photosynthesis. There is an additional cost of two ATPs associated with moving the three-carbon "ferry" molecule from the bundle sheath cell back to the mesophyll to pick up another molecule of atmospheric CO_2. Since the spatial separation in bundle-sheath cells minimizes O_2 concentrations in the locations where Rubisco is located, photorespiration is minimized (**Figure 24.3**). This arrangement of cells reduces photorespiration and increases the efficiency of photosynthesis for C_4 plants. In addition, C_4 plants require about half as much water as a C_3 plant. The reason C_4 plants require less water is due to the fact that the physical shape of the stomata and leaf structure of C_4 plants helps reduce water loss by developing a large CO_2 concentration gradient between the outside of the leaf (400 ppm) and the mesophyll cells (10 ppm). The large CO_2 concentration gradient reduces water loss via transpiration through the stomata.

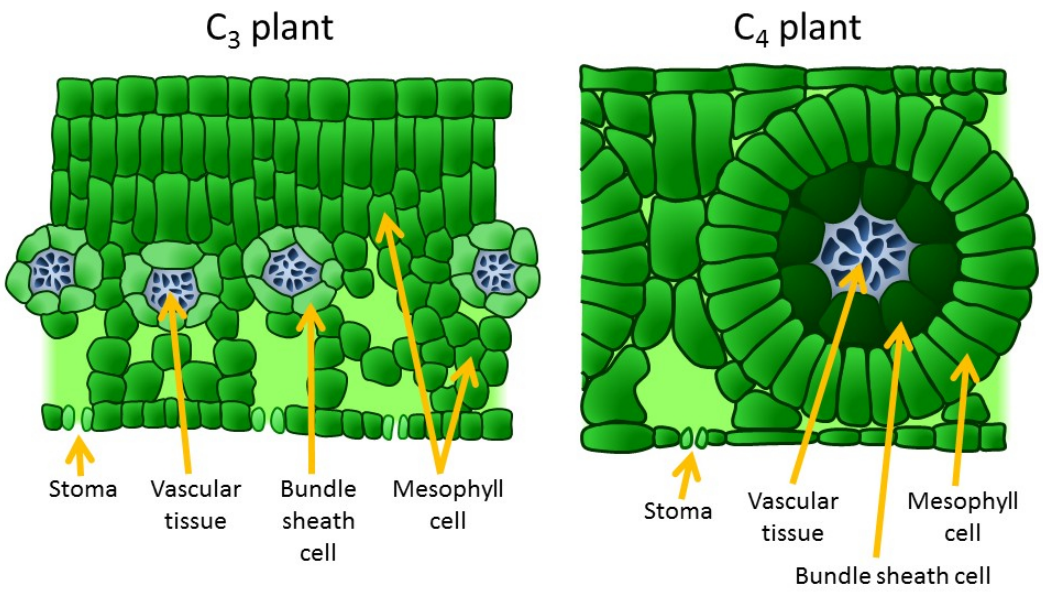


Figure 24.2 Cross section of a C_3 and C_4 plant leaf. Work by Eva Horne

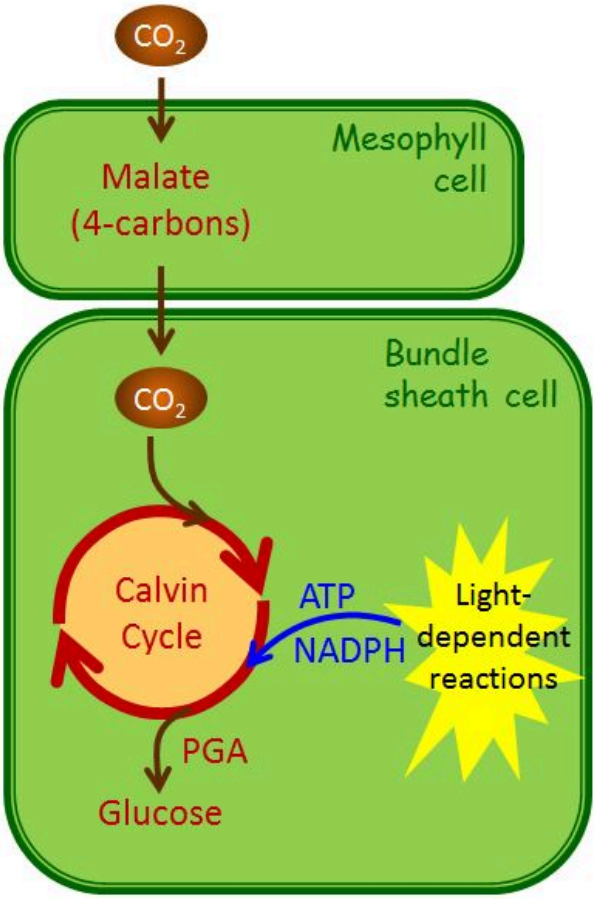


Figure 24.3 The spatial separation of Carbon fixation and the Calvin cycle in C_4 plants. Work by Eva Horne

The C_4 pathway is used in about 3% of all vascular plants; some examples are crabgrass, sugarcane and corn. C_4 plants are common in habitats that are hot, but are less abundant in areas that are cooler, because the enzyme that fixes the CO_2 in the mesophyll is less efficient at lower temperature. One hypothesis for the abundance of C_4 plants in hot habitats is that the benefits of reduced photorespiration and water loss exceeds the ATP cost of moving the the CO_2 from the mesophyll cell to bundle-sheath cell.

CAM

Many plants such as cacti and pineapples, which are adapted to arid environments, use a different energy and water saving pathway called crassulacean acid metabolism (CAM). This name comes from the family of plants (Crassulaceae) in which scientists first discovered the pathway. Instead of separating the light-dependent reactions and the use of CO_2 in the Calvin cycle spatially, CAM plants separate these processes temporally (**Figure 24.4**). At night, CAM plants open their stomata, and an enzyme in the mesophyll cells fix the CO_2 as an organic acid and store the organic acid in vacuoles until morning. During the day the light-dependent reactions supply the ATP and NADPH necessary for the Calvin cycle to function, and the CO_2 is released from those organic acids and used to make sugars. Plant species using CAM photosynthesis are the most water efficient of all; the stomata are only open at night when humidity is typically higher and the temperatures are much cooler (which serves to lower the diffusive gradient driving water loss from leaves). The CAM pathway is primarily an adaptation to water-limited environments; the fact that this pathway also stops photorespiration is an added benefit.

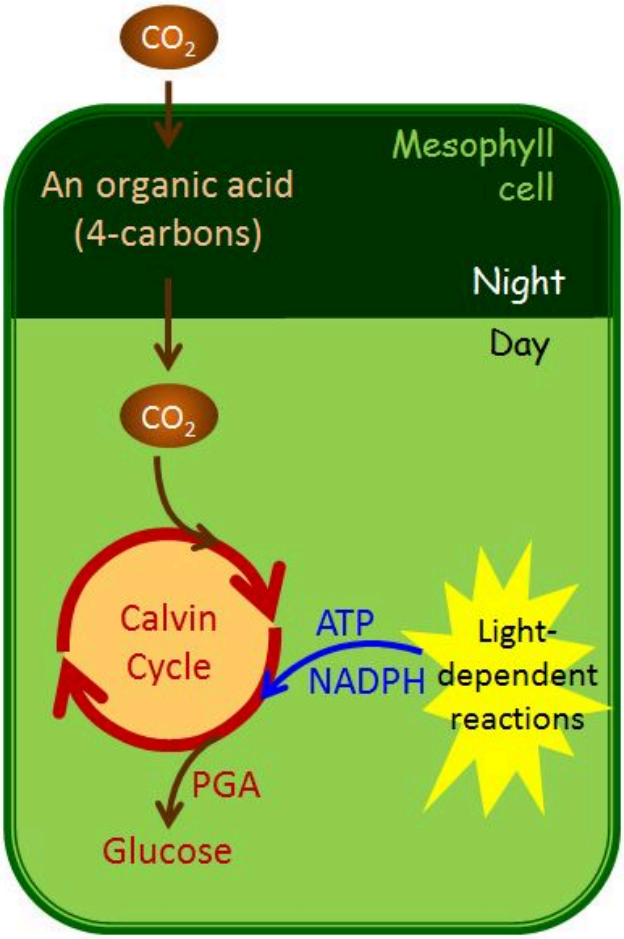


Figure 24.4 Temporal separation of Carbon fixation and the Calvin cycle in CAM plants. Work by Eva Horne

Overall, C_3, C_4 and CAM plants all use the Calvin cycle to make sugars from CO_2. However, the various ways in which plants fix CO_2 varies with the advantages and disadvantages associated with the mechanism and the habitats where plants can be found (Table 1).

As humans continue to burn fossil fuels, CO_2 levels in the atmosphere will continue to increase. This human alteration of the environment has sparked the development of a number of interesting questions. What influence will increasing CO_2 have on the distributions of C_3, C_4 and CAM plants? What influence will increasing CO_2 have on agricultural production? Is it possible that an increase in agricultural production by additional CO_2 in the atmosphere could offset or mitigate the decrease in agricultural production caused by climate change?

	C_3 plant	C_4 plant	CAM Plant
Cost	Photorespiration	ATP cost associated with fixing carbon twice. Carbon fixation is less efficient under cold conditions.	Reduced amount of fixed carbon, stomata only open at night
Benefits	Carbon fixation without using ATP	Reduced photorespiration and ability to fix Carbon under high temperatures and reduced water loss	Reduced photorespiration and reduced water loss
Separation of light-dependent reactions and carbon fixation	None, all of these reactions occur in the same cells	Spatial, these two sets of reactions occur in different cells	Temporal, these two sets of reactions occur at different times of day

Table 24.1 Characteristics of C_3, C_4 and CAM methods of fixing CO_2

Habitat	Cool and moist	Hot, not in cold environments (see cost.)	Hot and dry, large temperature differential between night and day

Table 24.1 Characteristics of C_3, C_4 and CAM methods of fixing CO_2

24.2 | Climate and the Effects of Global Climate Change

> "Every year, by burning fossil fuels, we release a million years of photosynthesis"

Daniel Nocera, an MIT scientist working to develop a process which uses sunlight and water to produce energy, quoted in "The Artificial Leaf" by Daniel Owen, *The New Yorker*, May 14, 2012.

All biomes are affected by global conditions, such as climate, that ultimately shape each biome's environment. Scientists who study climate have noted a series of marked changes that have gradually become increasingly evident during the last sixty years. **Global climate change** is the term used to describe altered global weather patterns, including a worldwide increase in temperature, due largely to rising levels of atmospheric carbon dioxide.

Climate and Weather

A common misconception about global climate change is that a specific weather event occurring in a particular region (for example, a very cool week in June in central Indiana) is evidence of global climate change. However, a cold week in June is a weather-related event and not a climate-related one. These misconceptions often arise because of confusion over the terms climate and weather.

Climate refers to the long-term, predictable atmospheric conditions of a specific area. The climate of a biome is characterized by having consistent temperature and annual rainfall ranges. Climate does not address the amount of rain that fell on one particular day in a biome or the colder-than-average temperatures that occurred on one day. In contrast, **weather** refers to the conditions of the atmosphere during a short period of time. Weather forecasts are usually made for 48-hour cycles. Long-range weather forecasts are available but can be unreliable.

To better understand the difference between climate and weather, imagine that you are planning an outdoor event in northern Wisconsin. You would be thinking about *climate* when you plan the event in the summer rather than the winter because you have long-term knowledge that any given Saturday in the months of May to August would be a better choice for an outdoor event in Wisconsin than any given Saturday in January. However, you cannot determine the specific day that the event should be held on because it is difficult to accurately predict the weather on a specific day. Climate can be considered "average" weather.

Global Climate Change

Climate change can be understood by approaching three areas of study:

- current and past global climate change
- causes of past and present-day global climate change
- ancient and current results of climate change

It is helpful to keep these three different aspects of climate change clearly separated when consuming media reports about global climate change. It is common for reports and discussions about global climate change to confuse the data showing that Earth's climate is changing with the factors that drive this climate change.

Evidence for Global Climate Change

Since scientists cannot go back in time to directly measure climatic variables, such as average temperature and precipitation, they must instead indirectly measure temperature. To do this, scientists rely on historical evidence of Earth's past climate.

Antarctic ice cores are a key example of such evidence. These ice cores are samples of polar ice obtained by means of drills that reach thousands of meters into ice sheets or high mountain glaciers. Viewing the ice cores is like traveling backwards through time; the deeper the sample, the earlier the time period. Trapped within the ice are bubbles of air and other biological evidence that can reveal temperature and carbon dioxide data. Antarctic ice cores have been collected and analyzed to indirectly estimate the temperature of the Earth over the past 400,000 years (**Figure 24.5a**). The 0 °C on this graph refers to the long-term average. Temperatures that are greater than 0 °C exceed Earth's long-term average temperature. Conversely, temperatures that are less than 0 °C are less than Earth's average temperature. This figure shows that there have been periodic cycles of increasing and decreasing temperature.

Before the late 1800s, the Earth has been as much as 9 °C cooler and about 3 °C warmer. Note that the graph in **Figure 24.5b** shows that the atmospheric concentration of carbon dioxide has also risen and fallen in periodic cycles; note the relationship between carbon dioxide concentration and temperature. **Figure 24.5b** shows that carbon dioxide levels in the atmosphere have historically cycled between 180 and 300 parts per million (ppm) by volume.

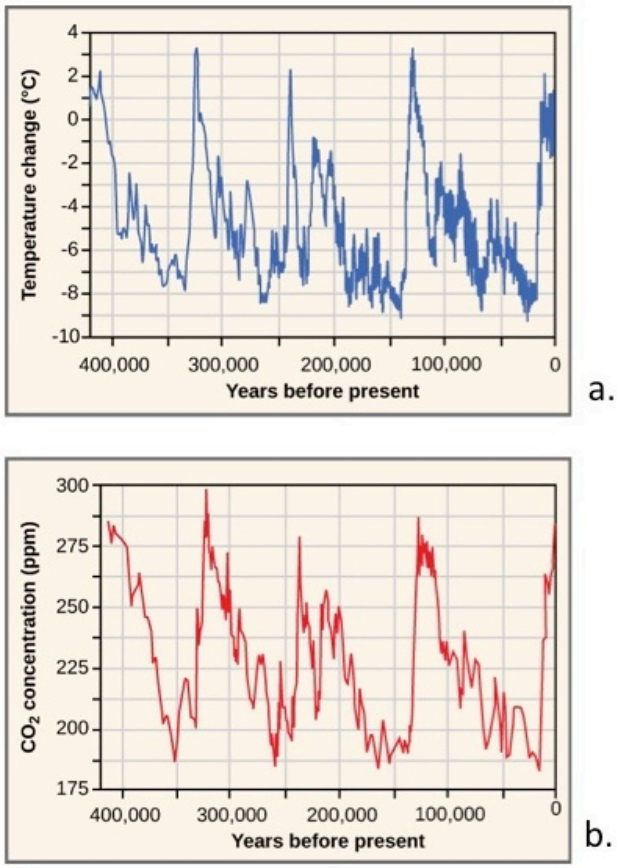


Figure 24.5 Ice at the Russian Vostok station in East Antarctica was laid down over the course 420,000 years and reached a depth of over 3,000 m. By measuring the amount of CO_2 trapped in the ice, scientists have determined past atmospheric CO_2 concentrations. Temperatures relative to modern day were determined from the amount of deuterium (an isotope of hydrogen) present.

Figure 24.5a does not show the last 2,000 years with enough detail to compare the changes of Earth's temperature during the last 400,000 years with the temperature change that has occurred in the more recent past. Two significant temperature anomalies, or irregularities, have occurred in the last 2000 years. These are the Medieval Climate Anomaly (or the Medieval Warm Period) and the Little Ice Age. A third temperature anomaly aligns with the Industrial Era. The Medieval Climate Anomaly occurred between 900 and 1300 AD. During this time period, many climate scientists think that slightly warmer weather conditions prevailed in many parts of the world; the higher-than-average temperature changes varied between 0.10 °C and 0.20 °C above the norm. Although 0.10 °C does not seem large enough to produce any noticeable change, it did free seas of ice. Because of this warming, the Vikings were able to colonize Greenland.

The Little Ice Age was a cold period that occurred between 1550 AD and 1850 AD. During this time, a slight cooling of a little less than 1 °C was observed in North America, Europe, and possibly other areas of the Earth. This 1 °C change in global temperature is a seemingly small deviation in temperature (as was observed during the Medieval Climate Anomaly);

however, it also resulted in noticeable changes. Historical accounts reveal a time of exceptionally harsh winters with much snow and frost.

The Industrial Revolution, which began around 1750, was characterized by changes in much of human society. Advances in agriculture increased the food supply, which improved the standard of living for people in Europe and the United States. New technologies were invented and provided jobs and cheaper goods. These new technologies were powered using fossil fuels, especially coal. The Industrial Revolution starting in the early nineteenth century ushered in the beginning of the Industrial Era. When a fossil fuel is burned, carbon dioxide is released. With the beginning of the Industrial Era, atmospheric carbon dioxide began to rise (**Figure 24.6**).

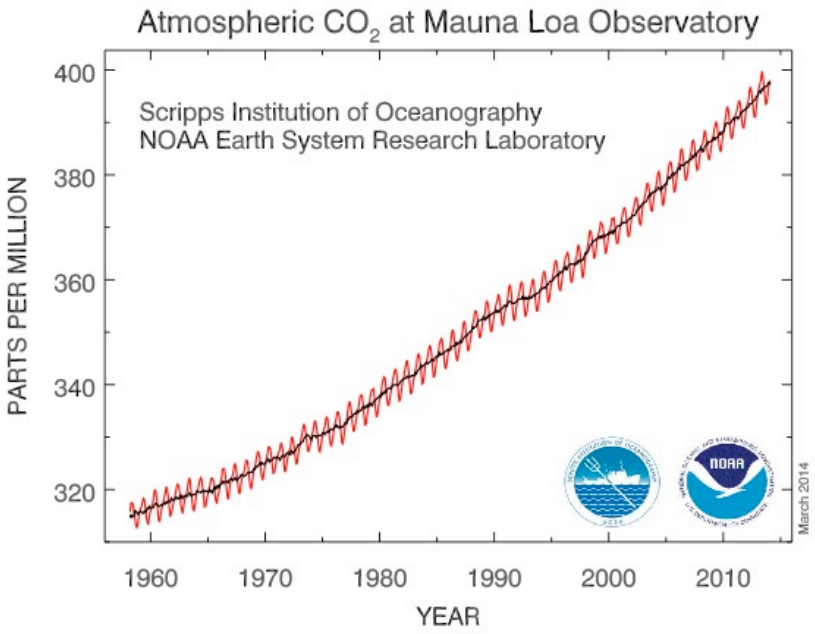


Figure 24.6 The atmospheric concentration of CO_2 has risen steadily since the beginning of industrialization. Graph from http://www.esrl.noaa.gov/gmd/webdata/ccgg/trends/co2_data_mlo.pdf

Current and Past Drivers of Global Climate Change

Scientists also use indirect evidence to determine the drivers, or factors, that may be responsible for climate change. The indirect evidence includes data collected using ice cores, boreholes (a narrow shaft bored into the ground), tree rings, glacier lengths, pollen remains, and ocean sediments. The data shows a correlation between the timing of temperature changes and drivers of climate change: before the Industrial Era (pre-1780), there were three drivers of climate change that were not related to human activity or atmospheric gases. The first of these is the Milankovitch cycles. The Milankovitch cycles describe the effects of slight changes in the Earth's orbit on Earth's climate. The length of the Milankovitch cycles ranges between 19,000 and 100,000 years. In other words, one could expect to see some predictable changes in the Earth's climate associated with changes in the Earth's orbit at a minimum of every 19,000 years.

The variation in the sun's intensity is the second natural factor responsible for climate change. Solar intensity is the amount of solar power or energy the sun emits in a given amount of time. There is a direct relationship between solar intensity and temperature. As solar intensity increases (or decreases), the Earth's temperature correspondingly increases (or decreases). Changes in solar intensity have been proposed as one of several possible explanations for the Little Ice Age.

Finally, volcanic eruptions are a third natural driver of climate change. Volcanic eruptions can last a few days, but the solids and gases released during an eruption can influence the climate over a period of a few years, causing short-term climate changes. The gases and solids released by volcanic eruptions can include carbon dioxide, water vapor, sulfur dioxide, hydrogen sulfide, hydrogen, and carbon monoxide. Generally, volcanic eruptions cool the climate. This occurred in 1783 when volcanos in Iceland erupted and caused the release of large volumes of sulfuric oxide. This led to haze-effect cooling, a global phenomenon that occurs when dust, ash, or other suspended particles block out sunlight and trigger lower global temperatures as a result; haze-effect cooling usually extends for one or more years. In Europe and North America, haze-

effect cooling produced some of the lowest average winter temperatures on record in 1783 and 1784.

Greenhouse gases are probably the most significant drivers of the climate. When heat energy from the sun strikes the Earth, gases known as **greenhouse gases** trap the heat in the atmosphere, as do the glass panes of a greenhouse keep heat from escaping. The greenhouse gases that affect Earth include carbon dioxide, methane, water vapor, nitrous oxide, and ozone. Approximately half of the radiation from the sun passes through these gases in the atmosphere and strikes the Earth. This radiation is converted into thermal radiation on the Earth's surface, and then a portion of that energy is re-radiated back into the atmosphere. Greenhouse gases, however, reflect much of the thermal energy back to the Earth's surface. The more greenhouse gases there are in the atmosphere, the more thermal energy is reflected back to the Earth's surface. Greenhouse gases absorb and emit radiation and are an important factor in the **greenhouse effect**: the warming of Earth due to carbon dioxide and other greenhouse gases in the atmosphere.

Evidence supports the relationship between atmospheric concentrations of carbon dioxide and temperature: as carbon dioxide rises, global temperature rises. Since 1950, the concentration of atmospheric carbon dioxide has increased from about 280 ppm to 382 ppm in 2006. In 2013, the atmospheric carbon dioxide concentration was 400 ppm.

Scientists look at patterns in data and try to explain differences or deviations from these patterns. The atmospheric carbon dioxide data reveal a historical pattern of carbon dioxide increasing and decreasing, cycling between a low of 180 ppm and a high of 300 ppm. Scientists have concluded that it took around 50,000 years for the atmospheric carbon dioxide level to increase from its low minimum concentration to its higher maximum concentration. However, starting recently, atmospheric carbon dioxide concentrations have increased beyond the historical maximum of 300 ppm. The current increases in atmospheric carbon dioxide have happened very quickly—in a matter of hundreds of years rather than thousands of years. What is the reason for this difference in the rate of change and the amount of increase in carbon dioxide? A key factor that must be recognized when comparing the historical data and the current data is the presence of modern human society; no other driver of climate change has yielded changes in atmospheric carbon dioxide levels at this rate or to this magnitude.

Human activity releases carbon dioxide and methane, two of the most important greenhouse gases, into the atmosphere in several ways. The primary mechanism that releases carbon dioxide is the burning of fossil fuels, such as gasoline, coal, and natural gas (**Figure 24.7**). Deforestation, cement manufacture, animal agriculture, the clearing of land, and the burning of forests are other human activities that release carbon dioxide. Methane (CH_4) is produced when bacteria break down organic matter under anaerobic conditions. Anaerobic conditions can happen when organic matter is trapped underwater (such as in rice paddies) or in the intestines of herbivores. Methane can also be released from natural gas fields and the decomposition that occurs in landfills. Another source of methane is the melting of clathrates. Clathrates are frozen chunks of ice and methane found at the bottom of the ocean. When water warms, these chunks of ice melt and methane is released. As the ocean's water temperature increases, the rate at which clathrates melt is increasing, releasing even more methane. This leads to increased levels of methane in the atmosphere, which further accelerates the rate of global warming. This is an example of the positive feedback loop that is leading to the rapid rate of increase of global temperatures.

Figure 24.7 The burning of fossil fuels in industry and by vehicles releases carbon dioxide and other greenhouse gases into the atmosphere. (credit: "Pöllö"/Wikimedia Commons)

Documented Results of Climate Change: Past and Present

Scientists have geological evidence of the consequences of long-ago climate change. Modern-day phenomena such as retreating glaciers and melting polar ice cause a continual rise in sea level. Meanwhile, changes in climate can negatively affect organisms.

Past Climate Change

Global warming has been associated with at least one planet-wide extinction event during the geological past. The Permian extinction event occurred about 251 million years ago toward the end of the roughly 50-million-year-long geological time span known as the Permian period. This geologic time period was one of the three warmest periods in Earth's geologic history. Scientists estimate that approximately 70 percent of the terrestrial plant and animal species and 84 percent of marine species became extinct, vanishing forever near the end of the Permian period. Organisms that had adapted to wet and warm climatic conditions, such as annual rainfall of 300–400 cm (118–157 in) and 20 °C–30 °C (68 °F–86 °F) in the tropical wet forest, may not have been able to survive the Permian climate change.

Present Climate Change

A number of global events have occurred that may be attributed to climate change during our lifetimes. Glacier National Park in Montana is undergoing the retreat of many of its glaciers, a phenomenon known as glacier recession. In 1850, the area contained approximately 150 glaciers. By 2010, however, the park contained only about 24 glaciers greater than 25 acres in size. One of these glaciers is the Grinnell Glacier (**Figure 24.8**) at Mount Gould. Between 1966 and 2005, the size of Grinnell Glacier shrank by 40 percent. Similarly, the mass of the ice sheets in Greenland and the Antarctic is decreasing: Greenland lost 150–250 km^3 of ice per year between 2002 and 2006. In addition, the size and thickness of the Arctic sea ice is decreasing.

Figure 24.8 The effect of global warming can be seen in the continuing retreat of Grinnel Glacier. The mean annual temperature in the park has increased 1.33 °C since 1900. The loss of a glacier results in the loss of summer meltwaters, sharply reducing seasonal water supplies and severely affecting local ecosystems. (credit: modification of work by USGS)

This loss of ice is leading to increases in the global sea level. On average, the sea is rising at a rate of 1.8 mm per year. However, between 1993 and 2010 the rate of sea level increase ranged between 2.9 and 3.4 mm per year. A variety of factors affect the volume of water in the ocean, including the temperature of the water (the density of water is related to its temperature) and the amount of water found in rivers, lakes, glaciers, polar ice caps, and sea ice. As glaciers and polar ice caps melt, there is a significant contribution of liquid water that was previously frozen.

In addition to some abiotic conditions changing in response to climate change, many organisms are also being affected by the changes in temperature. Temperature and precipitation play key roles in determining the geographic distribution and phenology of plants and animals. (Phenology is the study of the effects of climatic conditions on the timing of periodic lifecycle events, such as flowering in plants or migration in birds.) Researchers have shown that 385 plant species in Great Britain are flowering 4.5 days sooner than was recorded earlier during the previous 40 years. In addition, insect-pollinated species were more likely to flower earlier than wind-pollinated species. The impact of changes in flowering date would be mitigated if the insect pollinators emerged earlier. This mismatched timing of plants and pollinators could result in injurious ecosystem effects because, for continued survival, insect-pollinated plants must flower when their pollinators are present.

24.3 | Human Population Continues to Grow

Introduction

> “Population, when unchecked, increases in a geometrical ratio. Subsistence increases only in an arithmetical ratio. A slight acquaintance with numbers will show the immensity of the first power in comparison of the second.”

Thomas Robert Malthus. *An Essay on the Principle of Population*, (1798)

Mathematical approaches to understanding animal population dynamics, which you learned about earlier in the semester, can be applied to human population growth. Malthus recognized long ago that exponential population growth might be a problem if resources increase non-exponentially. Resources can be increased by technology and other modifications to the environment. In fact, humans are not unique in their ability to alter their environment. For example, beaver dams alter the stream environment where they are built. Humans, however, have the ability to alter their environment to increase its carrying capacity sometimes to the detriment of other species (e.g., via artificial selection for crops that have a higher yield). Earth’s human population is growing rapidly, to the extent that some worry about the ability of the earth’s environment to sustain this population, as long-term exponential growth carries the potential risks of famine, disease, and large-scale death.

Although humans have increased the carrying capacity of their environment, the technologies used to achieve this

transformation have caused unprecedented changes to Earth's environment, altering ecosystems to the point where some may be in danger of collapse. The depletion of the ozone layer, erosion due to acid rain, and damage from global climate change are caused by human activities. The ultimate effect of these changes on our carrying capacity is unknown. As some point out, it is likely that the negative effects of increasing carrying capacity will outweigh the positive ones—the carrying capacity of the world for human beings might actually decrease.

The world's human population is currently experiencing exponential growth even though human reproduction is far below its biotic potential (**Figure 24.9**). To reach its biotic potential, all females would have to become pregnant every nine months or so during their reproductive years. Also, resources would have to be such that the environment would support such growth. Neither of these two conditions exists. In spite of this fact, human population is still growing exponentially.

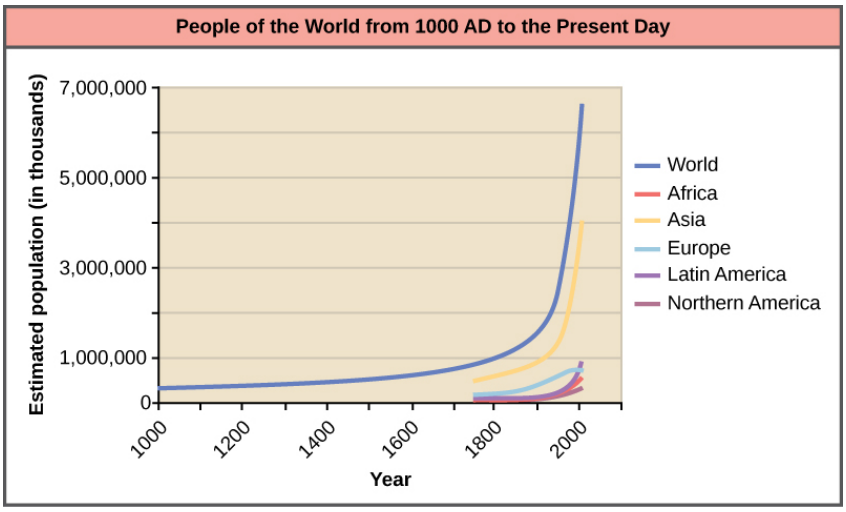


Figure 24.9 Human population growth since 1000 AD is exponential (dark blue line). Notice that while the population in Asia (yellow line), which has many economically underdeveloped countries, is increasing exponentially, the population in Europe (light blue line), where most of the countries are economically developed, is growing much more slowly.

A consequence of exponential human population growth is the time that it takes to add a particular number of humans to the Earth is becoming shorter. **Figure 24.10** shows that 123 years were necessary to add 1 billion humans in 1930, but it only took 24 years to add two billion people between 1975 and 1999. As already discussed, at some point it would appear that our ability to increase our carrying capacity indefinitely on a finite world is uncertain. Without new technological advances, the human growth rate has been predicted to slow in the coming decades. However, the population will still be increasing and the threat of overpopulation remains.

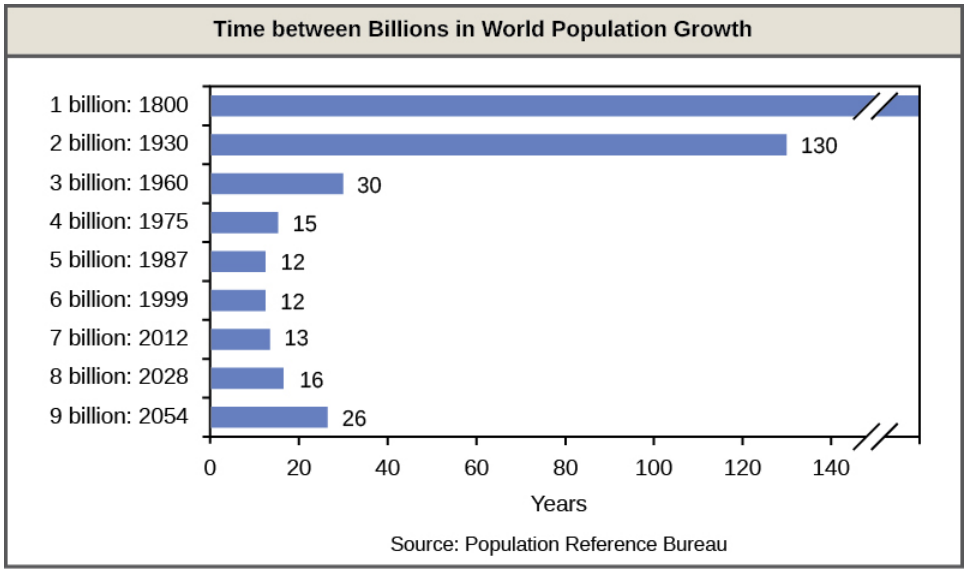


Figure 24.10 The time between the addition of each billion human beings to Earth decreases over time. (credit: modification of work by Ryan T. Cragun)

Overcoming Density-Dependent Regulation

Humans are unique in their ability to alter their environment with the conscious purpose of increasing its carrying capacity. This ability is a major factor responsible for human population growth and a way of overcoming density-dependent growth regulation. Much of this ability is related to human intelligence, society, and communication. Humans can construct shelter to protect them from the elements and have developed agriculture and domesticated animals to increase their food supplies. In addition, humans use language to communicate this technology to new generations, allowing them to improve upon previous accomplishments.

Other factors in human population growth are migration and public health. Humans originated in Africa, but have since migrated to nearly all inhabitable land on the Earth. Public health, sanitation, and the use of antibiotics and vaccines have decreased the ability of infectious disease to limit human population growth. In the past, diseases such as the bubonic plaque of the fourteenth century killed between 30 and 60 percent of Europe's population and reduced the overall world population by as many as 100 million people. Today, the threat of infectious disease, while not gone, is certainly less severe. According to the World Health Organization, global death from infectious disease declined from 14.2 million in 2000 to 9.9 million in 2011. To compare to some of the epidemics of the past, the percentage of the world's population killed between 1993 and 2002 decreased from 0.30 percent of the world's population to 0.24 percent. Thus, it appears that the influence of infectious disease on human population growth is becoming less significant.

Long-Term Consequences of Exponential Human Population Growth

Many dire predictions have been made about the world's population leading to a major crisis called the "population explosion." In the 1968 book *The Population Bomb*, biologist Dr. Paul R. Ehrlich wrote, "The battle to feed all of humanity is over. In the 1970s hundreds of millions of people will starve to death in spite of any crash programs embarked upon now. At this late date nothing can prevent a substantial increase in the world death rate."[1] While many critics view this statement as an exaggeration, the laws of exponential population growth are still in effect, and unchecked human population growth cannot continue indefinitely.

Efforts to control population growth led to the **one-child policy** in China, which used to include more severe consequences, but now imposes fines on urban couples who have more than one child. Due to the fact that some couples wish to have a male heir, many Chinese couples continue to have more than one child. The policy itself, its social impacts, and the effectiveness of limiting overall population growth are controversial. In spite of population control policies, the human population continues to grow. At some point the food supply may run out because of the subsequent need to produce more and more food to feed our population. The United Nations estimates that future world population growth may vary from 6 billion (a decrease) to 16 billion people by the year 2100. There is no way to know whether human population growth will moderate to the point where the crisis described by Dr. Ehrlich will be averted.

Another result of population growth is the endangerment of the natural environment. Many countries have attempted to reduce the human impact on climate change by reducing their emission of the greenhouse gas carbon dioxide. However, these treaties have not been ratified by every country, and many underdeveloped countries trying to improve their economic condition may be less likely to agree with such provisions if it means slower economic development. Furthermore, the role of human activity in causing climate change has become a hotly debated socio-political issue in some developed countries, including the United States. Thus, we enter the future with considerable uncertainty about our ability to curb human population growth and protect our environment.

1. Paul R. Erlich, prologue to *The Population Bomb*, (1968; repr., New York: Ballantine, 1970).

25 | INTRODUCTION TO ANIMALS

25.1 | Features of the Animal Kingdom

Introduction

> "Stones grow; plants grow and live; animals grow, live and feel."

Linneaus, in *Philosophia Botanica*, 1751

The differences between the animal, vegetable and mineral classes are a bit more complicated than the simple scheme of Linneaus. Members of the animal kingdom are incredibly diverse, but all animals share common features that distinguish them from organisms in other kingdoms. All animals are eukaryotic, multicellular organisms, and almost all animals have specialized tissues. All animals are motile, at least during certain life stages. Animals require a source of food to grow and develop. All animals are heterotrophic, ingesting living or dead organic matter. This form of obtaining energy distinguishes them from autotrophic organisms, such as most plants, which make their own nutrients through photosynthesis and from fungi that digest their food externally. Animals may be carnivores, herbivores, omnivores, or parasites (Figure 25.1). Most animals reproduce sexually: The offspring pass through a series of developmental stages that establish a determined body plan, unlike plants, for example, in which the exact shape of the body is indeterminate. The **body plan** refers to the shape of an animal.

(a) (b)

Figure 25.1 All animals that derive energy from food are heterotrophs. The (a) black bear is an omnivore, eating both plants and animals. The (b) heartworm *Dirofilaria immitis* is a parasite that derives energy from its hosts. It spends its larval stage in mosquitos and its adult stage infesting the hearts of dogs and other mammals, as shown here. (credit a: modification of work by USDA Forest Service; credit b: modification of work by Clyde Robinson)

Complex Tissue Structure

A hallmark trait of animals is specialized structures that are differentiated to perform unique functions. As multicellular organisms, most animals develop specialized cells that group together into tissues with specialized functions. A **tissue** is a collection of similar cells that had a common embryonic origin that share a specialized function. There are four main types of animal tissues: **nervous, muscle, connective, and epithelial**.

- Nervous tissue contains neurons, or nerve cells, which transmit nerve impulses.
- Muscle tissue contracts to cause all types of body movement from locomotion of the organism to movements within the body itself. You will learn more about the diferent types of muscle tissue later in this unit.
- Animals also have specialized connective tissues that provide many functions, including transport and structural support. Examples of connective tissues include blood and bone. Connective tissue is comprised of cells separated by extracellular material made of organic and inorganic materials, such as the protein and mineral deposits of bone.
- Epithelial tissue covers the internal and external surfaces of organs inside the animal body and the external surface of the body of the organism.

Animal Reproduction and Development

Most animals have diploid body (somatic) cells and a small number of haploid reproductive (gamete) cells produced through meiosis. Some exceptions exist: For example, in bees, wasps, and ants, the male is haploid because it develops from an unfertilized egg. Most animals undergo sexual reproduction, while many also have mechanisms of asexual reproduction.

Asexual Reproduction

Asexual reproduction, unlike sexual reproduction, produces offspring genetically identical to each other and to the parent. A number of animal species—especially those without backbones, but even some fish, amphibians, and reptiles—are capable of asexual reproduction. Asexual reproduction, except for occasional identical twinning, is absent in birds and mammals. The most common forms of asexual reproduction for stationary aquatic animals include budding and fragmentation, in which part of a parent individual can separate and grow into a new individual. In contrast, a form of asexual reproduction found in certain invertebrates and rare vertebrates is called parthenogenesis (or “virgin beginning”), in which unfertilized eggs develop into new offspring.

Sexual Reproduction and Embryonic Development

Almost all animal species are capable of reproducing sexually; for many, this is the only mode of reproduction possible. This distinguishes animals from fungi, protists, and bacteria, where asexual reproduction is common or exclusive. During sexual reproduction, the male and female gametes of a species combine in a process called fertilization. Typically, the small, motile male sperm travels to the much larger, sessile female egg. Sperm form is diverse and includes cells with flagella or amoeboid cells to facilitate motility. Fertilization and fusion of the gamete nuclei produce a zygote. Fertilization may be internal, especially in land animals, or external, as is common in many aquatic species.

After fertilization, a developmental sequence ensues as cells divide and differentiate. Many of the events in development are shared in groups of related animal species, and these events are one of the main ways scientists classify high-level groups of animals. During development, animal cells specialize and form tissues, determining their future morphology and physiology. In many animals, such as mammals, the young resemble the adult. Other animals, such as some insects and amphibians, undergo complete metamorphosis in which individuals enter one or more larval stages. For these animals, the young and the adult have different diets and sometimes habitats. In other species, a process of incomplete metamorphosis occurs in which the young somewhat resemble the adults and go through a series of stages separated by molts (shedding of the skin) until they reach the final adult form.

Features used in the Classification of Animals

Animals are classified according to morphological and developmental characteristics, such as a body plan. With the exception of sponges, the animal body plan is symmetrical. This means that their distribution of body parts is balanced along an axis. Additional characteristics that contribute to animal classification include the number of tissue layers formed during development, the presence or absence of an internal body cavity, and other features of embryological development.

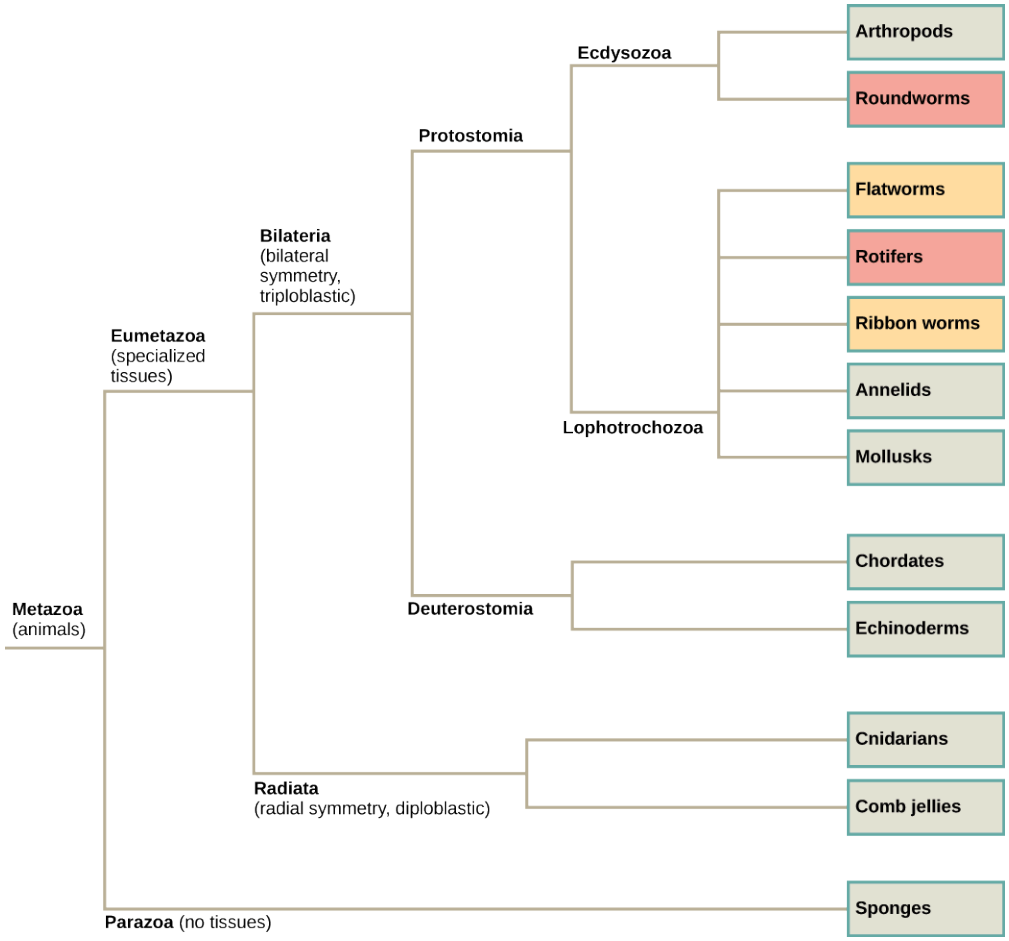


Figure 25.2 The phylogenetic tree of animals is based on morphological, fossil, and genetic evidence.

Body Symmetry

Animals may be asymmetrical, radial, or bilateral in form (**Figure 25.3**). **Asymmetrical** animals are animals with no pattern or symmetry; an example of an asymmetrical animal is a sponge (**Figure 25.3a**). An organism with **radial symmetry** (**Figure 25.3b**) has a longitudinal (up-and-down) orientation: Any plane cut along this up–down axis produces roughly mirror-image halves. An example of an organism with radial symmetry is a sea anemone.

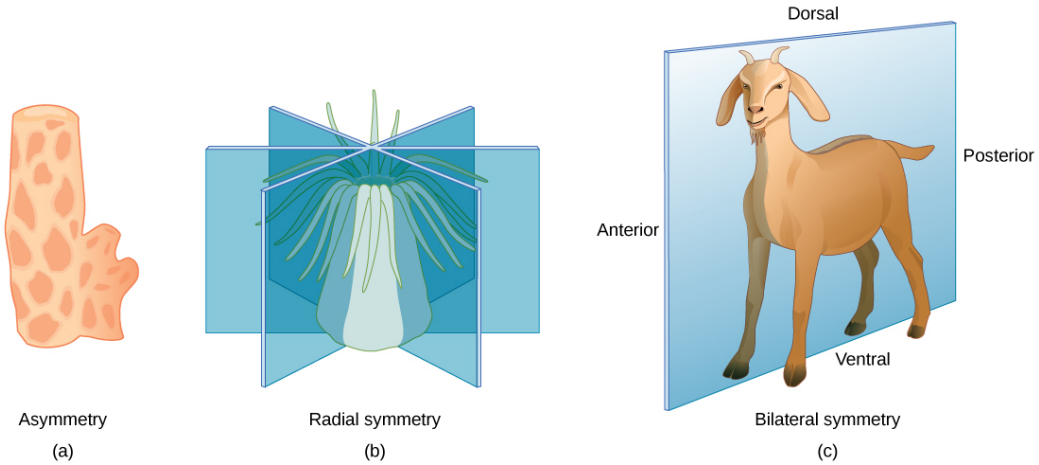


Figure 25.3 Animals exhibit different types of body symmetry. The (a) sponge is asymmetrical and has no planes of symmetry, the (b) sea anemone has radial symmetry with multiple planes of symmetry, and the (c) goat has bilateral symmetry with one plane of symmetry.

Bilateral symmetry is illustrated in **Figure 25.3c** using a goat. The goat also has upper and lower sides to it, but they are

not symmetrical. A vertical plane cut from front to back separates the animal into roughly mirror-image right and left sides. Animals with bilateral symmetry also have a “head” and “tail” (anterior versus posterior) and a back and underside (dorsal versus ventral).

Getting a Head

A product of bilateral symmetry and of organisms having a head (anterior) and a tail (posterior) is the fact organisms move through their environment head first. Since organisms move through their environment head first, a significant advantage can be seen with organisms that have a concentration of sensory organs in the head region or **cephalization**. In other words, placement of sensory organs around the mouth of a heterotroph allows it to be more efficient at finding food. In addition, organisms moving around with the sensory organs in the front will be more efficient at detecting potential predators and not becoming food.

Segmentation

The division of an animal into repeating body parts is called **segmentation**. You can clearly see segmentation in earthworms and millipeds, but in some insects and chordates the subdivisions are not as obvious. What are the advantages of a segmented body? Segmentation allows for greater flexibility and mobility. The repeating body parts allows for specialization of specific body parts, such as the development of legs, arm and wings.

Germ Layers

Nearly all animal species undergo a layering of early tissues during embryonic development. These layers are called **germ layers**. Each layer develops into a specific set of tissues and organs. Animals develop either two or three embryonic germ layers (Figure 25.4). The animals that display radial symmetry develop two germ layers, an inner layer (endoderm) and an outer layer (ectoderm). These animals are called diploblasts. Animals with bilateral symmetry develop three germ layers: an inner layer (endoderm), an outer layer (ectoderm), and a middle layer (mesoderm). Animals with three germ layers are called triploblasts.

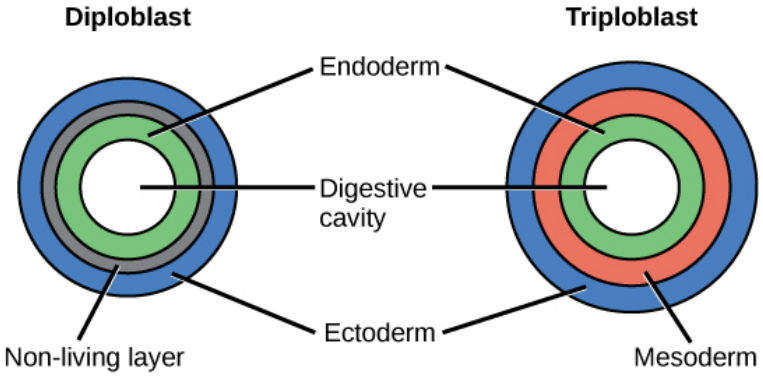


Figure 25.4 During embryogenesis, diploblasts develop two embryonic germ layers: an ectoderm and an endoderm. Triploblasts develop a third layer—the mesoderm—between the endoderm and ectoderm.

In triploblasts, the three germ layers develop in all the parts of an adult animal. The **endoderm** gives rise to the innermost lining of internal organs, such as those in the digestive tract, the liver, the pancreas and the lining of the lungs. The majority of organs and tissues in an adult animal such as the kidney, heart, muscles, blood vessels, bones and the dermis (inner layer of skin) develop from the **mesoderm**. Lastly, the **ectoderm** develops into the outermost layer of skin (epidermis), the lens and cornea of the eye, and the nervous system (brain and nerves).

Germ Layer	Tissue types
Ectoderm	Outer layer of skin, nerves, brain, cornea and lens of the eye
Mesoderm	Connective tissue of skin (dermis), bone, muscle (including cardiac muscle), cartilage, blood cells and blood vessels, fat cells, reproductive tract
Endoderm	Internal lining of organs of the digestive tract, internal lining of respiratory tract, liver

Table 25.1 Examples of tissue types that come from the three germ layers

Presence or Absence of a Coelom

Triploblasts may develop an internal body cavity lined with cells derived from mesoderm, called a **coelom** (pr. see-LŌM). This epithelial-lined cavity is a space, usually filled with fluid, which lies between the digestive system and the body wall. It houses organs such as the kidneys and spleen, and contains the circulatory system. An **Organ** is a differentiated structure that performs a specific function in an organism; it consists of many cells and of various tissue types. Triploblasts that do not develop a coelom are called **acoelomates**, and this internal region is completely filled with tissue, although they have a gut cavity. Examples of acoelomates include the flatworms. Animals with a true coelom are called **eucoelomates** (or coelomates) (**Figure 25.5**). A true coelom arises entirely within the mesoderm germ layer. Animals such as earthworms, snails, insects, starfish, and vertebrates are all eucoelomates. A third group of triploblasts has a body cavity that is derived partly from mesoderm and partly from endoderm tissue. These animals are called **pseudocoelomates**. Roundworms are examples of pseudocoelomates. New data on the relationships of pseudocoelomates suggest that these phyla are not closely related and so the evolution of the pseudocoelom must have occurred more than once (**Figure 25.2**). True coelomates can be further characterized based on features of their early embryological development.

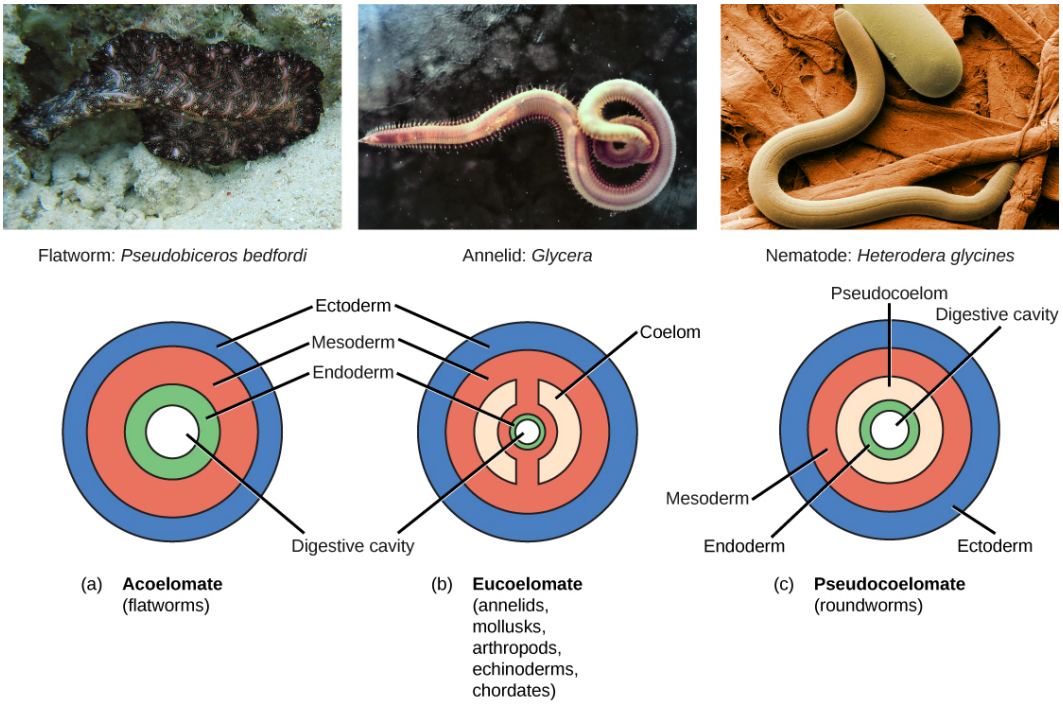


Figure 25.5 Triploblasts may be acoelomates, eucoelomates, or pseudocoelomates. Eucoelomates have a body cavity within the mesoderm, called a coelom, which is lined with mesoderm tissue. Pseudocoelomates have a similar body cavity, but it is lined with mesoderm and endoderm tissue. (credit a: modification of work by Jan Derk; credit b: modification of work by NOAA; credit c: modification of work by USDA, ARS)

Protostomes and Deuterostomes

Bilaterally symmetrical, triploblastic eucoelomates can be divided into two groups based on differences in their early embryonic development. You may have noticed in **Figure 25.2** the terms "Protostomia" and "Deuterostomia". Protostomes (members of the Protostomia) include phyla such as arthropods, mollusks, and annelids. Deuterostomes (members of the Deuterostomia) include the chordates and echinoderms. These two groups are named from which opening of the digestive cavity develops first: mouth or anus. The word *protostome* comes from Greek words meaning “mouth first,” and *deuterostome* originates from words meaning “mouth second” (in this case, the anus develops first). This difference reflects the fate of a structure called the blastopore (**Figure 25.6**), which becomes the mouth in protostomes and the anus in deuterostomes. There are other developmental differences between protostomes and deuterostomes, including the mode of formation of the coelom and the pattern of early cell division of the embryo.

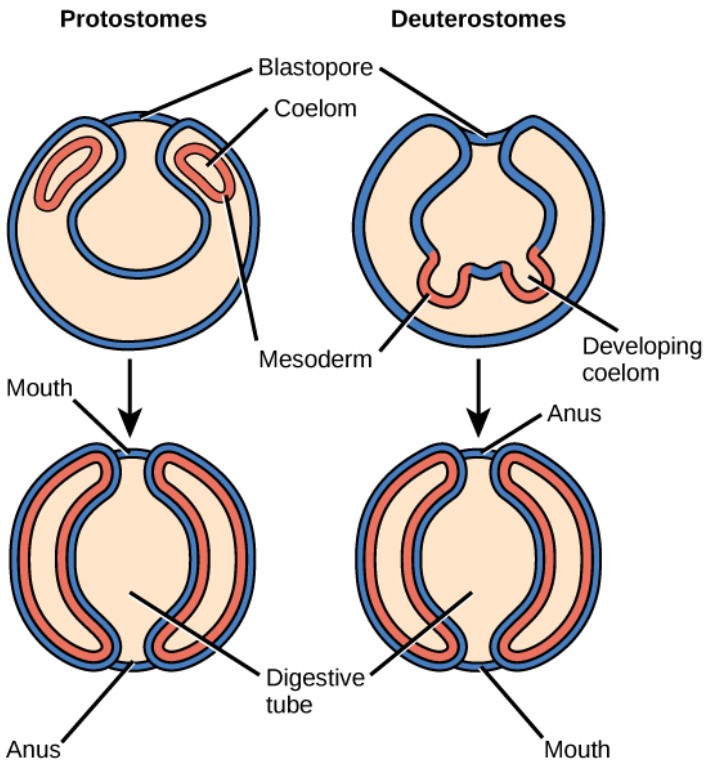


Figure 25.6 Eucoelomates can be divided into two groups, protostomes and deuterostomes, based on their early embryonic development. Two of these differences include the origin of the mouth opening and the way in which the coelom is formed.

25.2 | Animal Tissue Types

Introduction

> “ The elementary parts of all tissues are formed of cells in an analogous, though very diversified manner, so that it may be asserted, that there is one universal principle of development for the elementary parts of organisms, however different, and that this principle is the formation of cells.”

Theodor Schwann, in *Microscopic Researches into the Accordance in the Structure and Growth of Animals and Plants*, 1839

Cells make tissues, tissues make organs, and organs make organ systems. The tissues of multicellular, complex animals are four primary types: epithelial, connective, muscle, and nervous. Recall that tissues are groups of similar cells carrying out a specific related functions. Many different tissues combine to form an **organ**—like the skin or kidney—that have specialized functions within the body. Organs are organized into organ systems to perform functions; examples include the circulatory system, which consists of the heart and blood vessels, and the digestive system, consisting of several organs, including the stomach, intestines, liver, and pancreas. Organ systems come together to create an entire organism.

Epithelial Tissues

Epithelial tissues cover the outside of organs and structures in the body and line the lumens of organs in a single layer or multiple layers of cells. The types of epithelia are classified by the shapes of cells present and the number of layers of

cells. Epithelia composed of a single layer of cells is called **simple epithelia**; epithelial tissue composed of multiple layers is called **stratified epithelia**. Table 25.2 summarizes the different types of epithelial tissues.

Different Types of Epithelial Tissues

Cell shape	Description	Location
squamous	flat, irregular round shape	simple: lung alveoli, capillaries stratified: skin, mouth, vagina
cuboidal	cube shaped, central nucleus	glands, renal tubules
columnar	tall, narrow, nucleus toward base tall, narrow, nucleus along cell	simple: digestive tract pseudostratified: respiratory tract
transitional	round, simple but appear stratified	urinary bladder

Table 25.2

Squamous Epithelia

Squamous epithelial cells are generally round, flat, and have a small, centrally located nucleus. The cell outline is slightly irregular, and cells fit together to form a covering or lining. When the cells are arranged in a single layer (simple epithelia), they facilitate diffusion in tissues, such as the areas of gas exchange in the lungs and the exchange of nutrients and waste at blood capillaries.

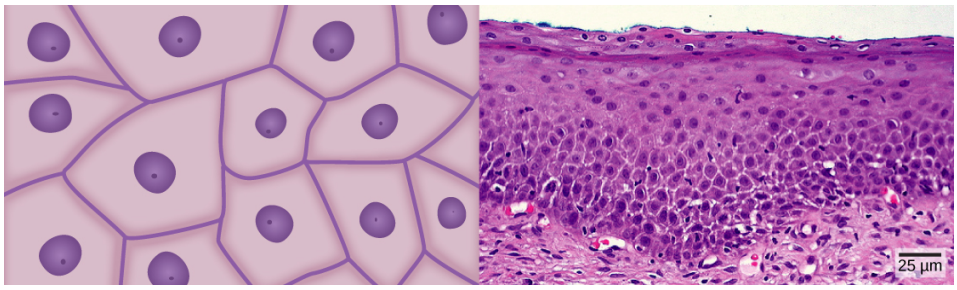


Figure 25.7 Squamous epithelia cells (a) have a slightly irregular shape, and a small, centrally located nucleus. These cells can be stratified into layers, as in (b) this human cervix specimen. (credit b: modification of work by Ed Uthman; scale-bar data from Matt Russell)

Figure 25.7**a** illustrates a layer of squamous cells with their membranes joined together to form an epithelium. Image Figure 25.7**b** illustrates squamous epithelial cells arranged in stratified layers, where protection is needed on the body from outside abrasion and damage. This is called a stratified squamous epithelium and occurs in the skin and in tissues lining the mouth and vagina.

Cuboidal Epithelia

Cuboidal epithelial cells, shown in Figure 25.8, are cube-shaped with a single, central nucleus. They are most commonly found in a single layer representing a simple epithelia in glandular tissues (e.g. pancreas, mammary gland, etc.). They are also found in the walls of tubules and in the ducts of the kidney and liver.

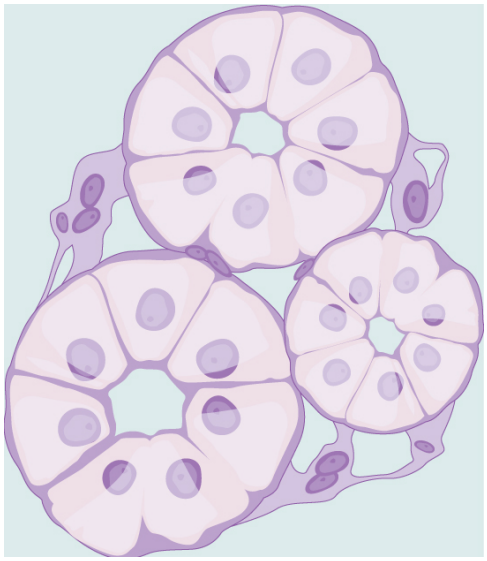


Figure 25.8 Simple cuboidal epithelial cells line tubules in the mammalian kidney, where they are involved in filtering the blood.

Columnar Epithelia

Columnar epithelial cells are taller than they are wide: they resemble a stack of columns in an epithelial layer, and are most commonly found in a single-layer arrangement. The nuclei of columnar epithelial cells in the digestive tract appear to be lined up at the base of the cells, as illustrated in **Figure 25.9**. These cells absorb material from the lumen of the digestive tract and prepare it for entry into the body through the circulatory and lymphatic systems.

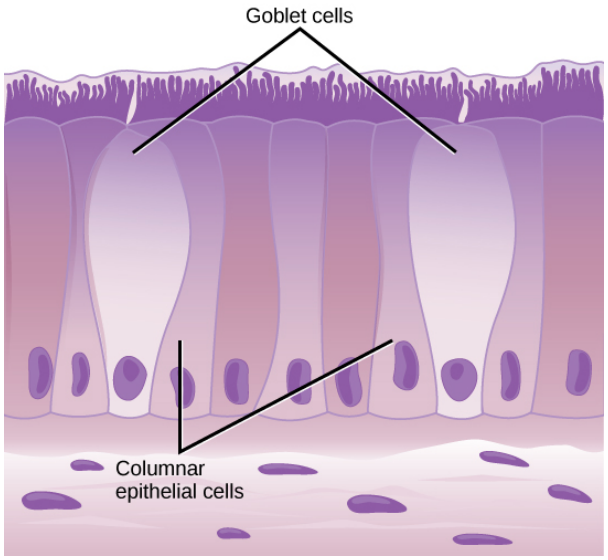


Figure 25.9 Simple columnar epithelial cells absorb material from the digestive tract. Goblet cells secret mucous into the digestive tract lumen (at top of the figure).

Columnar epithelial cells lining the respiratory tract appear to be stratified. However, each cell is attached to the base membrane of the tissue and, therefore, they are simple tissues. The nuclei are arranged at different levels in the layer of cells, making it appear as though there is more than one layer, as seen in **Figure 25.10**. This is called pseudostratified, columnar epithelia. This cellular covering has cilia at the apical, or free, surface of the cells. The cilia enhance the movement of mucous and trapped particles out of the respiratory tract, helping to protect the system from invasive microorganisms and harmful material that has been breathed into the body. Goblet cells are interspersed in some tissues (such as the lining of the trachea). The goblet cells contain mucous that traps irritants, which in the case of the trachea keep these irritants from getting into the lungs.

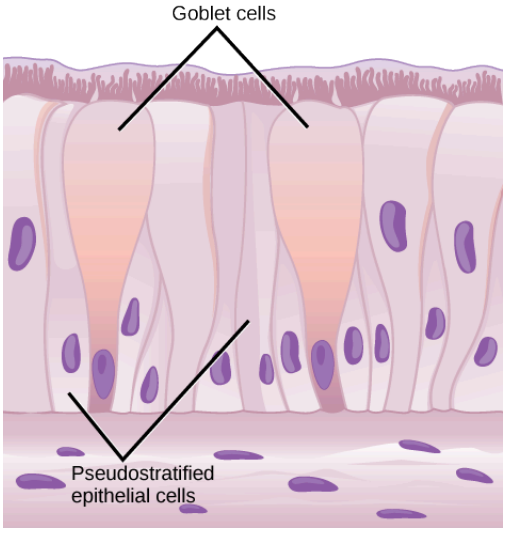


Figure 25.10 Pseudostratified columnar epithelia line the respiratory tract. They exist in one layer, but the arrangement of nuclei at different levels makes it appear that there is more than one layer. Goblet cells interspersed between the columnar epithelial cells secrete mucous into the respiratory tract (at top of the figure).

Connective Tissues

Connective tissues are made up of a matrix consisting of living cells and a non-living substance, called the ground substance. The ground substance is made of an organic substance (usually a protein) and an inorganic substance (usually a mineral or water). The principal cell of connective tissues is the fibroblast. This cell makes the fibers found in nearly all of the connective tissues. Fibroblasts are motile, able to carry out mitosis, and can synthesize whichever connective tissue is needed. Macrophages, lymphocytes, and, occasionally, leukocytes can be found in some connective tissues. Other cells that can be found in different connective tissues provide functions that are important for those specific tissues. For example, osteocytes (found in bone) and adipocytes (found in fat) are specialized cells that are critical for the formation and function of those tissues.

The organic portion or protein fibers found in connective tissues are either collagen, elastic, or reticular fibers. Collagen fibers provide strength to the tissue, preventing it from being torn or separated from the surrounding tissues. Elastic fibers are made of the protein elastin; this fiber can stretch to one and one half of its length and return to its original size and shape. Elastic fibers provide flexibility to the tissues. Reticular fibers are the third type of protein fiber found in connective tissues. This fiber consists of thin strands of collagen that form a network of fibers to support the tissue and other organs to which it is connected. The various types of connective tissues, the types of cells and fibers they are made of, and sample locations of the tissues is summarized in **Table 25.3**.

Connective Tissues

Tissue	Cells	Fibers	Location
loose/areolar	fibroblasts, macrophages, some lymphocytes, some neutrophils	few: collagen, elastic, reticular	around blood vessels; anchors epithelia
dense, fibrous connective tissue	fibroblasts, macrophages,	mostly collagen	irregular: skin regular: tendons, ligaments
cartilage	chondrocytes, chondroblasts	hyaline: few collagen fibrocartilage: large amount of collagen	shark skeleton, fetal bones, human ears, intervertebral discs
bone	osteoblasts, osteocytes, osteoclasts	some: collagen, elastic	vertebrate skeletons

Table 25.3

Connective Tissues

Tissue	Cells	Fibers	Location
adipose	adipocytes	few	adipose (fat)
blood	red blood cells, white blood cells	none	blood

Table 25.3

Loose/Areolar Connective Tissue

Loose connective tissue, also called areolar connective tissue, has a sampling of all of the components of a connective tissue. As illustrated in **Figure 25.11**, loose connective tissue has some fibroblasts; macrophages are present as well. In this figure, collagen fibers are relatively wide and stain a light pink, while elastic fibers are thin and stain dark blue to black. The space between the formed elements of the tissue is filled with the matrix. The material in the connective tissue gives it a loose consistency similar to a cotton ball that has been pulled apart. Loose connective tissue is found around every blood vessel and helps to keep the vessel in place. The tissue is also found around and between most body organs. In summary, areolar tissue is tough, yet flexible.

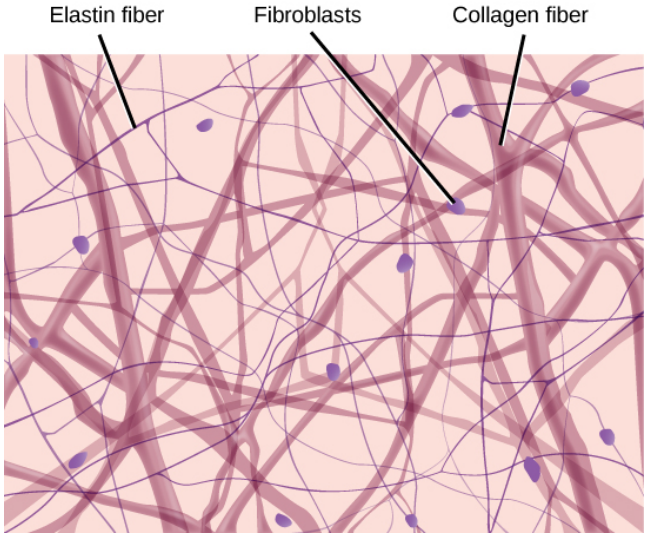


Figure 25.11 Loose connective tissue is composed of loosely woven collagen and elastic fibers. The fibers and other components of the connective tissue matrix are secreted by fibroblasts.

Fibrous Connective Tissue

Fibrous connective tissues contain large amounts of collagen fibers and few cells or matrix material. The fibers can be arranged irregularly or regularly with the strands lined up in parallel. Irregularly arranged fibrous connective tissues are found in areas of the body where stress occurs from all directions, such as the dermis of the skin. Regular fibrous connective tissue, shown in **Figure 25.12**, is found in tendons (which connect muscles to bones) and ligaments (which connect bones to bones).

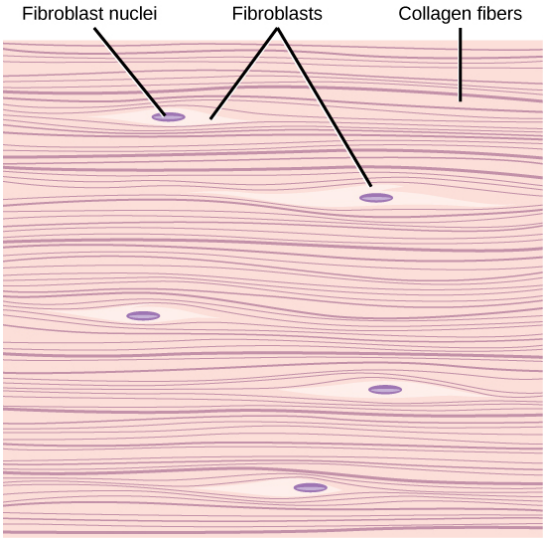


Figure 25.12 Fibrous connective tissue from the tendon has strands of collagen fibers lined up in parallel.

Cartilage

Cartilage is a connective tissue with a large amount of extracellular matrix and variable amounts of fibers. The cells, called chondrocytes, make the matrix and fibers of the tissue. Chondrocytes are found in spaces within the tissue (lacunae).

A cartilage with few collagen and elastic fibers is hyaline cartilage, illustrated in **Figure 25.13**. The lacunae are randomly scattered throughout the tissue and the matrix takes on a milky or scrubbed appearance with routine histological stains. Sharks have cartilaginous skeletons, as does nearly the entire human skeleton during a specific pre-birth developmental stage. A remnant of this cartilage persists in the outer portion of the human nose. Hyaline cartilage is also found at the ends of long bones, reducing friction and cushioning the articulations of these bones.

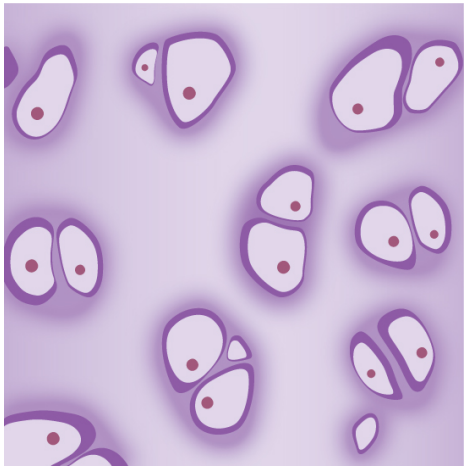


Figure 25.13 Hyaline cartilage consists of a matrix with cells called chondrocytes embedded in it. The chondrocytes exist in cavities in the matrix called lacunae.

Elastic cartilage has a large amount of elastic fibers, giving it tremendous flexibility. The ears of most vertebrate animals contain this cartilage as do portions of the larynx, or voice box. Fibrocartilage contains a large amount of collagen fibers, giving the tissue tremendous strength. Fibrocartilage comprises the intervertebral discs in vertebrate animals. Hyaline cartilage found in movable joints such as the knee and shoulder becomes damaged as a result of age or trauma. Damaged hyaline cartilage is replaced by fibrocartilage and results in the joints becoming “stiff.”

Bone

Bone, or osseous tissue, is a connective tissue that has a large amount of two different types of matrix material. The organic matrix is similar to the matrix material found in other connective tissues, including some amount of collagen and elastic fibers. This gives strength and flexibility to the tissue. The inorganic matrix consists of mineral salts—mostly calcium phosphate—that give the tissue hardness. In most organisms, bone is being remodeled by the action of cells called osteoclasts and osteoblasts (see below).

There are three types of cells in bone: osteoblasts, osteocytes, and osteoclasts. Osteoblasts are active in making bone for growth and remodeling. Osteoblasts deposit bone material into the matrix and, after the matrix surrounds them, they continue to live, but in a reduced metabolic state as osteocytes. Osteocytes are found in lacunae of the bone. Osteoclasts are active in breaking down bone for bone remodeling, and they provide access to calcium stored in tissues. Osteoclasts are usually found on the surface of the tissue.

Bone can be divided into two types: compact and spongy. Compact bone is found in the shaft (or diaphysis) of a long bone and the surface of the flat bones, while spongy bone is found in the end (or epiphysis) of a long bone. Compact bone is organized into subunits called **osteons**, as illustrated in Figure 25.14. A blood vessel and a nerve are found in the center of the structure within the Haversian canal, with radiating circles of lacunae around it known as lamellae. The wavy lines seen between the lacunae are microchannels called canaliculi; they connect the lacunae to aid diffusion between the cells. Spongy bone is made of tiny plates called trabeculae these plates serve as struts to give the spongy bone strength. Over time, these plates can break causing the bone to become less resilient.

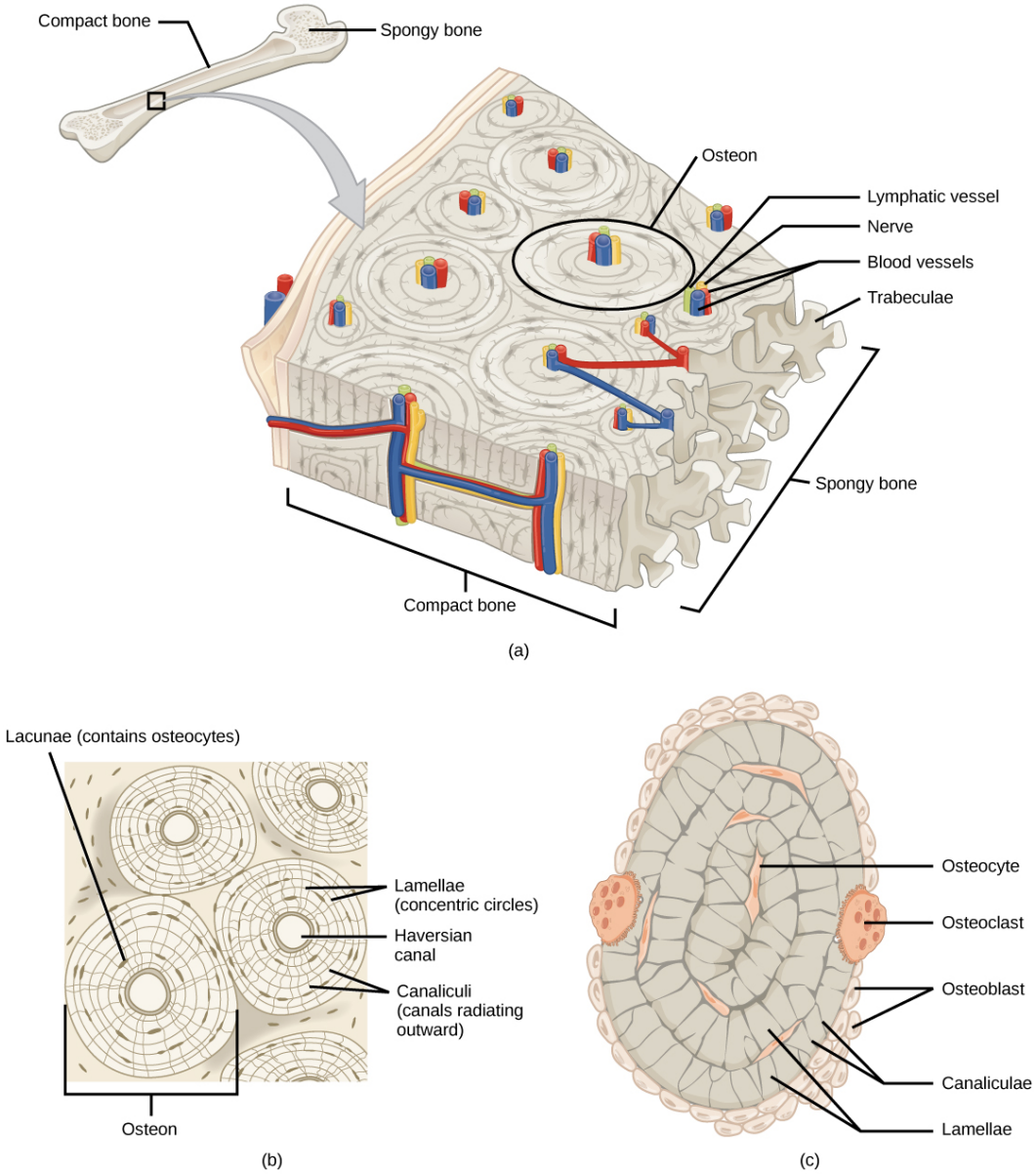


Figure 25.14 (a) Compact bone is a dense matrix on the outer surface of bone. Spongy bone, inside the compact bone, is porous with web-like trabeculae. (b) Compact bone is organized into rings called osteons. Blood vessels, nerves, and lymphatic vessels are found in the central Haversian canal. Rings of lamellae surround the Haversian canal. Between the lamellae are cavities called lacunae. Canaliculi are microchannels connecting the lacunae together. (c) Osteoblasts surround the exterior of the bone. Osteoclasts bore tunnels into the bone and osteocytes are found in the lacunae.

Adipose Tissue

Adipose tissue, or fat tissue, is considered a connective tissue even though it does not have fibroblasts or a real matrix, and only has a few fibers. Adipose tissue is made up of cells called adipocytes (descendants of fibroblasts) that collect and store fat in the form of triglycerides. Adipose tissues serve as energy stores, and additionally serve as insulation to help maintain body temperatures, allowing animals to be endothermic. They also function as cushions to prevent damage to body organs. Under a microscope, adipose tissue cells appear empty due to the extraction of fat during the processing of the material for viewing, as seen in **Figure 25.15**. The thin lines in the image are the plasma membranes, and the nuclei are the small, black dots at the edges of the cells.

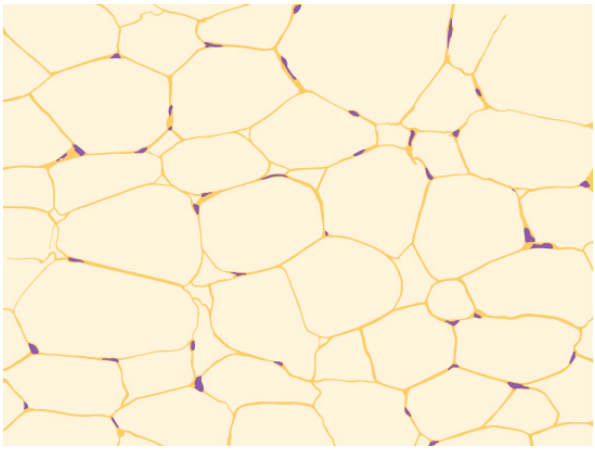


Figure 25.15 Adipose is a connective tissue is made up of cells called adipocytes. Adipocytes have small nuclei localized at the cell edge. The interior of these large cells is filled with triglycerides, commonly known as "fat".

Blood

Blood is considered a connective tissue because it has a fluid matrix (plasma) and is derived from the germ layer known as mesoderm. . The living cell types are red blood cells (RBC), also called erythrocytes, and white blood cells (WBC), also called leukocytes (**Figure 25.16**).

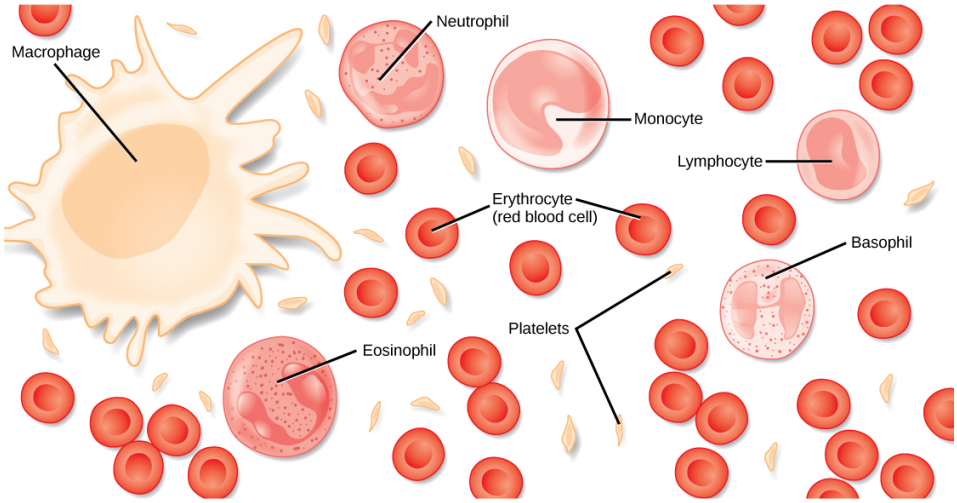


Figure 25.16 Blood is a connective tissue that has a fluid matrix, called plasma, and no fibers. Erythrocytes (red blood cells), the predominant cell type, are involved in the transport of oxygen and carbon dioxide. Also present are various leukocytes (white blood cells) involved in immune response.

The cell found in greatest abundance in blood is the erythrocyte, or red blood cell ("erythro" = red). The principal function of an erythrocyte is to carry and deliver oxygen to the tissues. There are millions of erythrocytes in every milliliter of your blood. Mammalian erythrocytes lose their nuclei and mitochondria when they mature and are released from the bone marrow where they are generated. Fish, amphibian, and avian red blood cells maintain their nuclei and mitochondria throughout the cell's life.

Leukocytes, or white blood cells ("leuko" = white), are the other main cellular component of blood. There are 5,000-10,000 leukocytes in every milliliter of your blood. These include cells called lymphocytes, as well as neutrophils, monocytes, and

others. Lymphocytes function primarily in the immune response to foreign antigens or material, which you will learn in more detail later in this unit. Neutrophils are phagocytic (they engulf other cells or objects and digest them) cells, and they participate in one of the early lines of defense against microbial invaders or fungal invaders. Another leukocyte that is found in the peripheral blood is the monocyte. Monocytes give rise to phagocytic macrophages that clean up dead and damaged cells in the body, whether they are foreign or from the host animal. Two additional leukocytes in the blood are eosinophils and basophils—both help to facilitate the inflammatory response.

The slightly granular material among the cells in **Figure 25.16** are cytoplasmic fragments of cells formed in the bone marrow. These are called platelets or thrombocytes. Platelets are a key player in the formation of blood clots, which are important in keeping your blood in your body when you get cut or scraped.

Muscle Tissues

There are three types of muscle in animal bodies: smooth, skeletal, and cardiac. They differ by the presence or absence of striations or bands, the number and location of nuclei, whether they are voluntarily or involuntarily controlled, and their location within the body. **Table 25.4** summarizes these differences. You will learn more about these cells and their functions later in this unit

Types of Muscles

Type of Muscle	Striations	Nuclei	Control	Location
smooth	no	single, in center	involuntary	visceral organs
skeletal	yes	many, at periphery	voluntary	skeletal muscles
cardiac	yes	single, in center	involuntary	heart

Table 25.4

Smooth Muscle

Smooth muscle does not have striations in its cells. It has a single, centrally located nucleus, as shown in **Figure 25.17**. Constriction of smooth muscle occurs under involuntary, autonomic nervous control and in response to local conditions in the tissues. Smooth muscle tissue is also called non-striated as it lacks the banded appearance of skeletal and cardiac muscle. The walls of blood vessels, the tubes of the digestive system, and the tubes of the reproductive systems are composed of mostly smooth muscle.

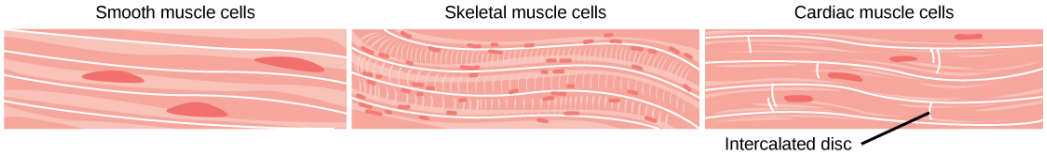


Figure 25.17 Smooth muscle cells do not have striations, while skeletal muscle cells do. Cardiac muscle cells have striations, but, unlike the multinucleate skeletal cells, they have only one nucleus. Cardiac muscle tissue also has intercalated discs, specialized regions running along the plasma membrane that join adjacent cardiac muscle cells and assist in passing an electrical impulse from cell to cell.

Skeletal Muscle

Skeletal muscle has striations across its cells caused by the arrangement of the contractile proteins actin and myosin. These muscle cells are relatively long and have multiple nuclei along the edge of the cell. Skeletal muscle is under voluntary, somatic nervous system control and is found in the muscles that move bones. **Figure 25.17** illustrates the histology of skeletal muscle.

Cardiac Muscle

Cardiac muscle, shown in **Figure 25.17**, is found only in the heart. Like skeletal muscle, it has cross striations in its cells, but cardiac muscle has a single, centrally located nucleus. Cardiac muscle is not under voluntary control but can be influenced by the autonomic nervous system to speed up or slow down. A structure found only in cardiac muscle cells is is at the end of the cell where it abuts the next cardiac cell in the row. This structure is called an intercalated disc: it assists in passing electrical impulses efficiently from one cell to the next, and maintains the strong synchrony needed to make the heart chambers work together efficiently.

Nervous Tissues

Nervous tissues are made of cells specialized to receive and transmit electrical impulses from specific areas of the body and to send them to specific locations in the body. The main cell of the nervous system is the neuron, illustrated in **Figure 25.18**. The large structure with a central nucleus is the cell body of the neuron. Projections from the cell body are either dendrites specialized in receiving input or a single axon specialized in transmitting impulses. Some glial cells are also shown. Astrocytes regulate the chemical environment of the nerve cell, and oligodendrocytes insulate the axon so the electrical nerve impulse is transferred more efficiently. Other glial cells that are not shown support the nutritional and waste requirements of the neuron. Some of the glial cells are phagocytic and remove debris or damaged cells from the tissue. A nerve in your body consists of neurons and glial cells.

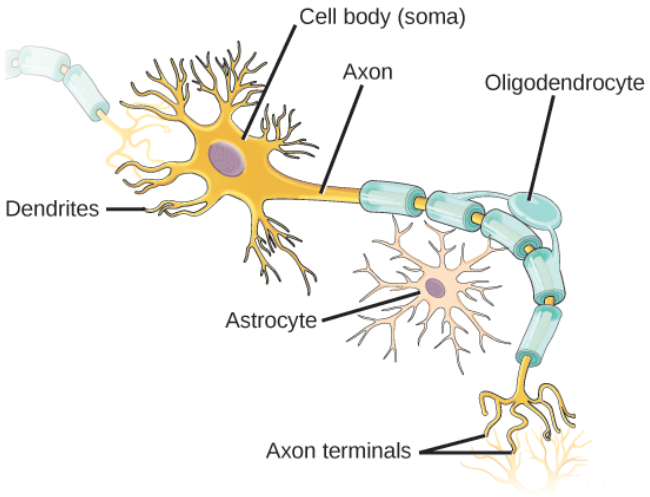


Figure 25.18 The neuron has projections called dendrites that receive signals and projections called axons that send signals. Also shown are two types of glial cells: astrocytes regulate the chemical environment of the nerve cell, and oligodendrocytes insulate the axon so the electrical nerve impulse is transferred more efficiently.

25.3 | Sponges and Cnidarians

Introduction

> “Jellyfish are 97% water or something, so how much are they doing? Just give them another 3% and make them water. It's more useful.”

Karl Pilkington, *Handslapped by a Jellyfish*, 2007

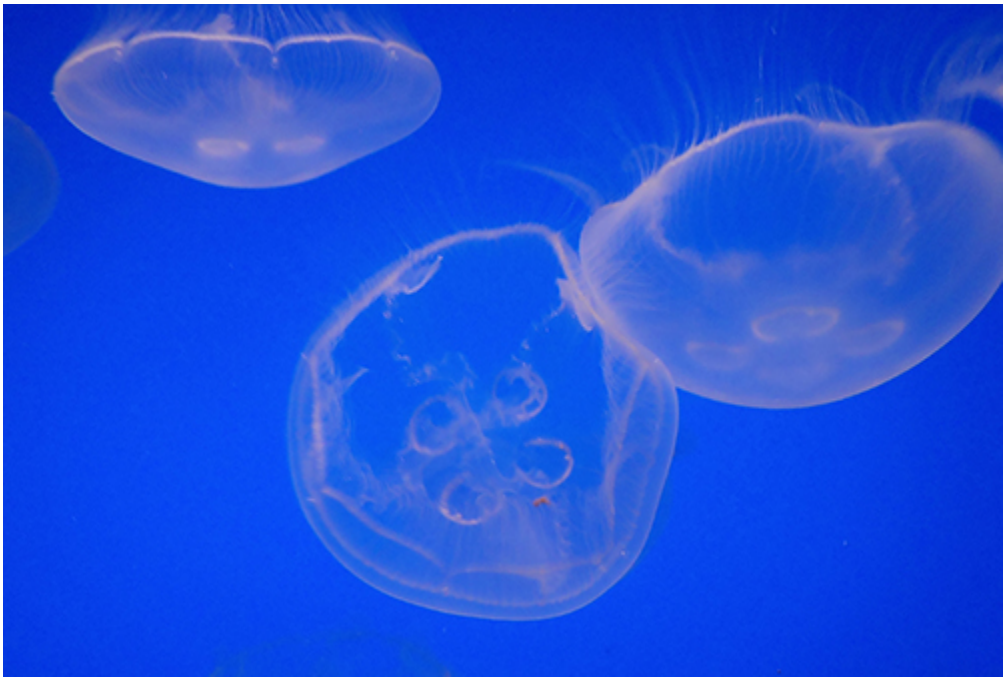


Figure 25.19 Jellyfish Jellyfish in the Monterey Aquarium, Monterey California. Image courtesy of David A. Rintoul

Pilkington's disdain for the jellyfish (a member of the group of animals we call Cnidarians) is misplaced. Jellyfish, besides being spectacular and fun to watch, also provide food for other marine animals, including sea turtles. These simple animals have been around for a long time, and can also teach us lots of lessons about animals and animal evolution.

The kingdom of animals is informally divided into invertebrate animals, those without a backbone, and vertebrate animals, those with a backbone. Although in general we are most familiar with vertebrate animals, the vast majority of animal species, about 95 percent, are invertebrates. Invertebrates include millions of species in about 32 phyla, and we will only hit the highlights in the subsequent sections of this text.

The sponges and the cnidarians represent the simplest of animals. Sponges appear to represent an early stage of multicellularity in the animal clade. Although they have specialized cells for particular functions, they **lack true tissues** in which specialized cells are organized into functional groups. Sponges are similar to what might have been the ancestor of animals: a colonial, flagellated protist. The cnidarians, or the jellyfish and their kin, are the simplest animal group that displays true tissues, although they possess only two tissue layers.

Sponges

Animals in subkingdom Parazoa represent the simplest animals and include the sponges, or phylum **Porifera** (**Figure 25.20**). All sponges are aquatic and the majority of species are marine. Sponges live in intimate contact with water, which plays a role in their feeding, gas exchange, and excretion. Much of the body structure of the sponge is dedicated to moving water through the body so it can filter out food, absorb dissolved oxygen, and eliminate wastes.

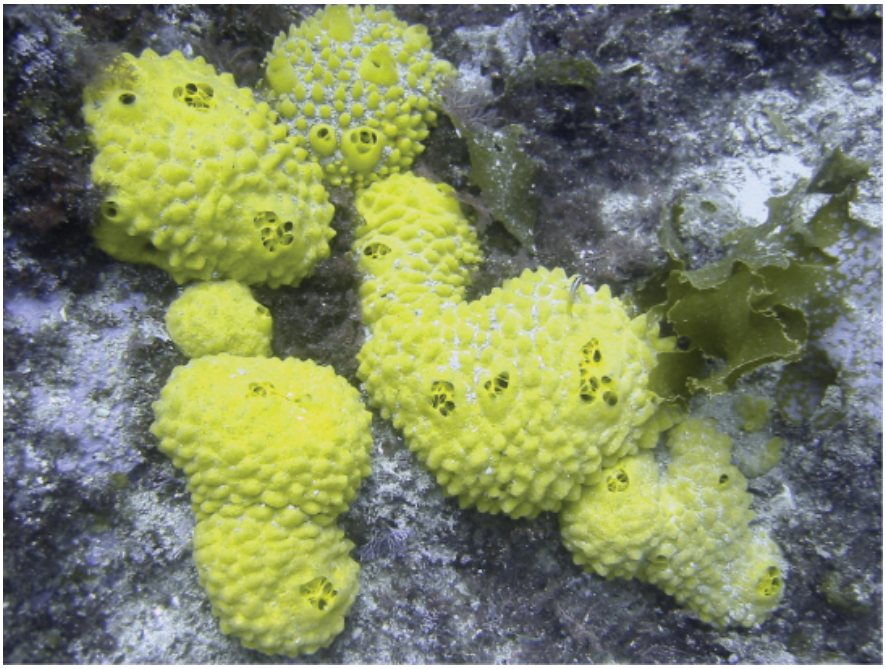
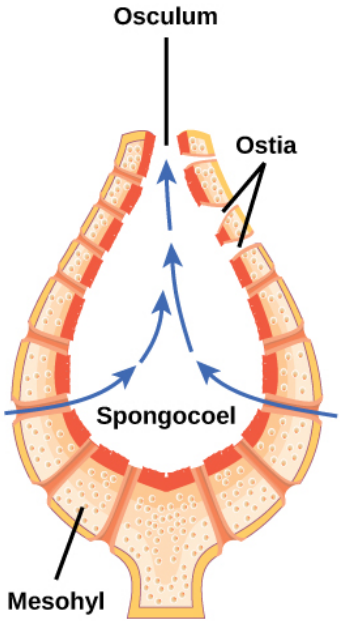


Figure 25.20 Sponges are members of the phylum Porifera, which contains the simplest animals. (credit: Andrew Turner)

The body of the simplest sponges takes the shape of a cylinder with a large central cavity, the spongocoel. Water enters the spongocoel from numerous pores in the body wall. Water flows out through a large opening called the osculum (**Figure 25.21**). However, sponges exhibit a diversity of body forms, which vary in the size and branching of the spongocoel, the number of osculi, and where the cells that filter food from the water are located.

Sponges consist of an outer layer of flattened cells and an inner layer of cells called choanocytes separated by a jelly-like substance called mesohyl. The mesohyl contains embedded amoeboid cells that secrete tiny needles called spicules or protein fibers that help give the sponge its structural strength. The cell body of the choanocyte is embedded in mesohyl but protruding into the spongocoel is a mesh-like collar surrounding a single flagellum. The beating of flagella from all choanocytes moves water through the sponge. Food particles are trapped in mucus produced by the sieve-like collar of the choanocytes and are ingested by phagocytosis and digested within those cells. Amoebocytes take up nutrients repackaged in food vacuoles of the choanocytes and deliver them to other cells within the sponge.



Basic sponge body plan

Figure 25.21 The sponge's basic body plan is shown.

Reproduction in Sponges

Despite their lack of complexity, sponges are clearly successful organisms, having persisted on Earth for more than half a billion years. Lacking a true digestive system, sponges depend on the intracellular digestive processes of their choanocytes for their energy intake. The limit of this type of digestion is that food particles must be smaller than individual cells. Gas exchange, circulation, and excretion occur by diffusion between cells and the water.

Sponges reproduce both sexually and asexually. Asexual reproduction is either by fragmentation (in which a piece of the sponge breaks off and develops into a new individual), or budding (an outgrowth from the parent that eventually detaches). But sponges are also capable of producing gametes, although both types of gametes can be produced in the same individual (hermaphroditism). Sponges may be sequentially hermaphroditic, producing eggs first and sperm later. Eggs areise from amoebocytes and are retained within the spongocoel, whereas sperm arise from choanocytes and a ejected via the osculum. These sperm are carried by the water and fertilize the eggs of other sponges. Larval development starts withing the sponge, and free-swimming larvae are then released via the osculum. This is the only time that sponges exhibit one of the hallmarks of the animal kingdom, motility. The larvae then attach to a substrate and spend their adult lives in the same spot.

Cnidarians

The phylum **Cnidaria** includes animals that show radial symmetry and are diploblastic (have two germ layers instead of the three that most animals have). Nearly all (about 99 percent) cnidarians are marine species, but there are freshwater jellyfish, even in Kansas! Cnidarians have specialized cells known as cnidocytes ("stinging cells") containing organelles called nematocysts. These cells are concentrated around the mouth and tentacles of the animal and can immobilize prey with toxins. Nematocysts contain coiled threads that may bear barbs. The outer wall of the cell has a hairlike projection that is sensitive to touch. When touched, the cells fire the toxin-containing coiled threads that can penetrate and stun the predator or prey (see **Figure 25.22**).

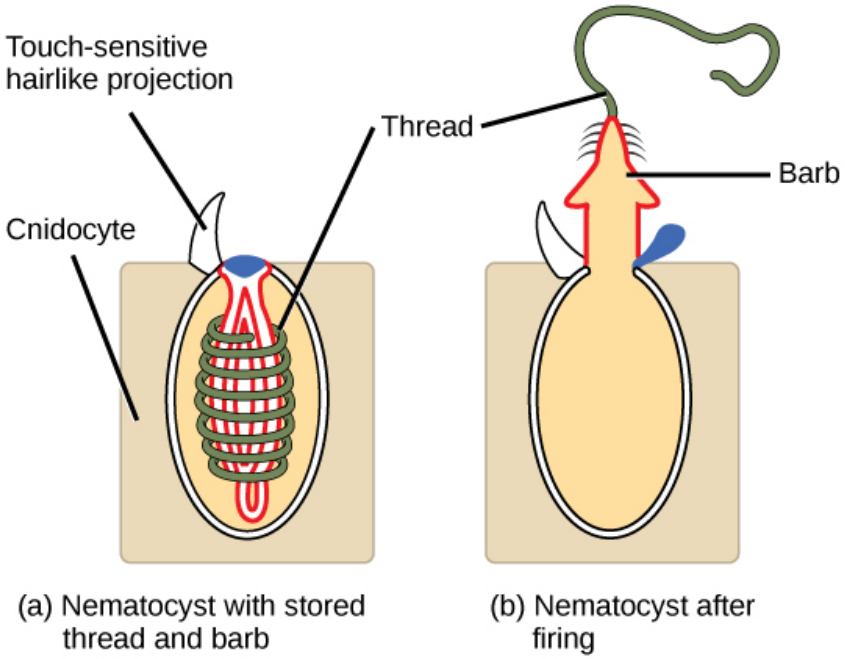


Figure 25.22 Animals from the phylum Cnidaria have stinging cells called cnidocytes. Cnidocytes contain large organelles called (a) nematocysts that store a coiled thread and barb. When hairlike projections on the cell surface are touched, (b) the thread, barb, and a toxin are fired from the organelle.

Cnidarians display two distinct body plans: polyp or "stalk" and medusa or "bell" (**Figure 25.23**). Examples of the polyp form are freshwater species of the genus *Hydra*; perhaps the best-known medusoid animals are the jellies (jellyfish). Polyps are sessile as adults, with a single opening to the digestive system (the mouth) facing up with tentacles surrounding it. Medusae are motile, with the mouth and tentacles hanging from the bell-shaped body. In other cnidarians, both a polyp and medusa form exist, and the life cycle alternates between these forms.

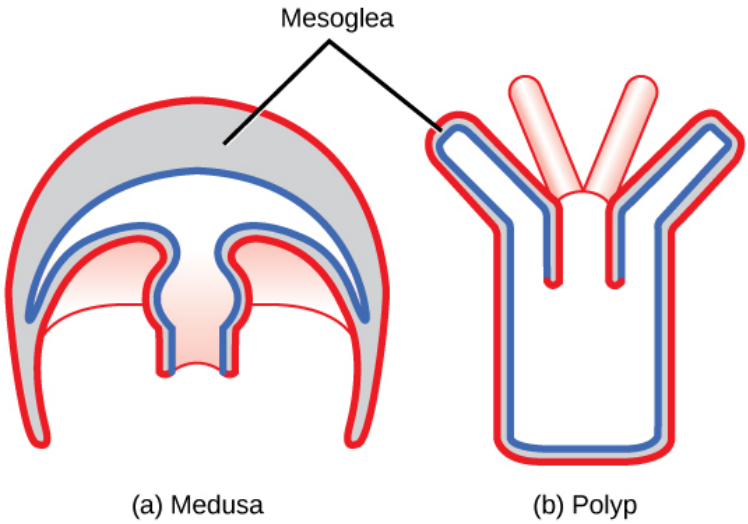


Figure 25.23 Cnidarians have two distinct body plans, the (a) medusa and the (b) polyp. All cnidarians have two tissue layers, with a jelly-like mesoglea between them.

Physiological Processes of Cnidarians

All cnidarians have two tissue layers. The outer layer is called the epidermis, whereas the inner layer is called the gastrodermis and lines the digestive cavity. Between these two layers is a non-living, jelly-like mesoglea. There are differentiated cell types in each tissue layer, such as nerve cells, enzyme-secreting cells, and nutrient-absorbing cells, as well as intercellular connections between the cells. However, organs and organ systems are not present in this phylum.

The nervous system is primitive, with nerve cells scattered across the body in a network. The function of the nerve cells is

to carry signals from sensory cells and to contractile cells. Groups of cells in the nerve net form nerve cords that may be essential for more rapid transmission. Cnidarians perform extracellular digestion, with digestion completed by intracellular digestive processes. Food is taken into the gastrovascular cavity, enzymes are secreted into the cavity, and the cells lining the cavity absorb the nutrient products of the extracellular digestive process. The gastrovascular cavity has only one opening that serves as both a mouth and an anus (an incomplete digestive system). Like the sponges, Cnidarian cells exchange oxygen, carbon dioxide, and nitrogenous wastes by diffusion between cells in the epidermis and gastrodermis with water.

25.4 | Flatworms, Nematodes, and Arthropods

Introduction

> “In the last ten years we have come to realize humans are more like worms than we ever imagined.”

Bruce Alberts, American scientist and editor of *Science* magazine

The use of a simple nematode, *Caenorhabditis elegans*, as a model system for developmental biology has indeed revealed many similarities between that simple creature and ourselves. The animal phyla of this and subsequent modules are **triploblastic** (i.e., have THREE primary germ layers, ectoderm, endoderm and mesoderm in the embryo). They have an embryonic mesoderm sandwiched between the ectoderm and endoderm. Most of these phyla are also bilaterally symmetrical, meaning that a longitudinal section will divide them into right and left sides that are mirror images of each other. Associated with bilateralism is the beginning of **cephalization**, the evolution of a concentration of nervous tissues and sensory organs in the head of the organism, which is where the organism first encounters its environment.

The flatworms are acoelomate organisms that include free-living and parasitic forms. The nematodes, or roundworms, possess a pseudocoelom and consist of both free-living and parasitic forms. Finally, the arthropods, one of the most successful taxonomic groups on the planet, are coelomate organisms with a hard exoskeleton and jointed appendages.

Flatworms

Physiological Processes of Flatworms

Most flatworms are parasitic, including important parasites of humans. Free-living species of flatworms are predators or scavengers, whereas parasitic forms feed from the tissues of their hosts. Digestion is extracellular, with enzymes secreted into the gut interior by cells lining the tract, and digested materials taken into the same cells by phagocytosis. One group, the cestodes, does not have a digestive system, because their parasitic lifestyle and the environment in which they live (suspended within the digestive cavity of their host) allows them to absorb nutrients directly across their body wall. Flatworms have an excretory system with a network of tubules throughout the body that open to the environment and nearby flame cells, whose cilia beat to direct waste fluids concentrated in the tubules out of the body. The system is responsible for regulation of dissolved salts and excretion of nitrogenous wastes. The nervous system consists of a pair of nerve cords running the length of the body with connections between them and a large ganglion or concentration of nerve cells at the anterior end of the worm; here, there may also be a concentration of photosensory and chemosensory cells (**Figure 25.24**).

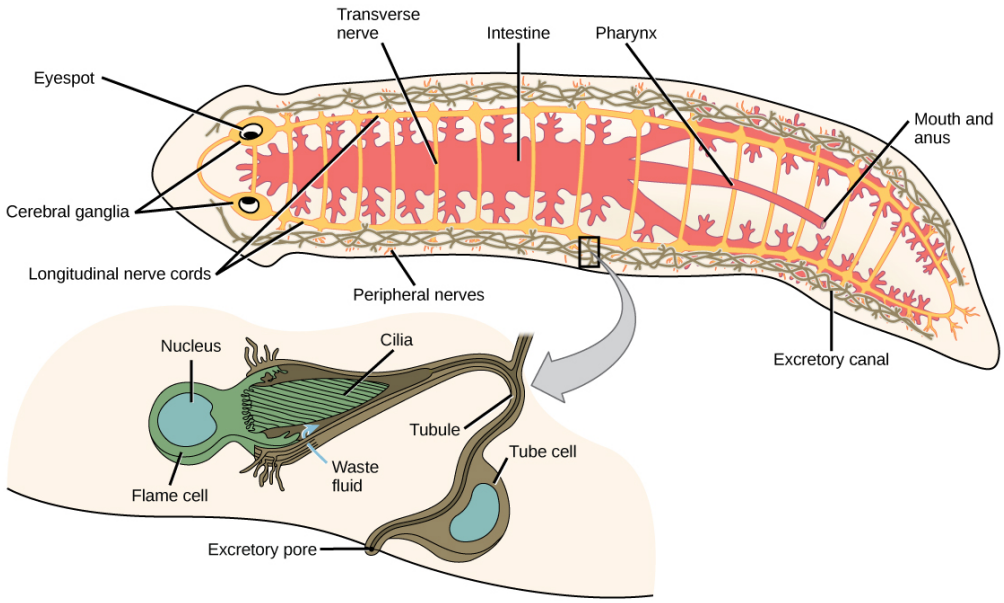


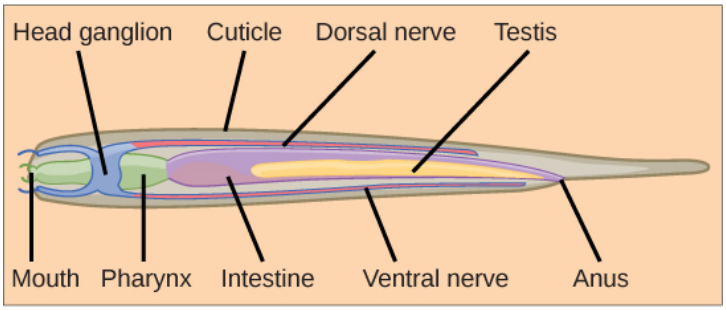
Figure 25.24 This planarian is a free-living flatworm that has an incomplete digestive system, an excretory system with a network of tubules throughout the body, and a nervous system made up of nerve cords running the length of the body with a concentration of nerves and photosensory and chemosensory cells at the anterior end.

Since there is no circulatory or respiratory system, gas and nutrient exchange is dependent on diffusion and intercellular junctions. This necessarily limits the thickness of the body in these organisms, constraining them to be "flat" worms. Most flatworm species are monoecious (hermaphroditic, possessing both sets of sex organs), and fertilization is typically internal. Asexual reproduction is common in some groups in which an entire organism can be regenerated from just a part of itself.

Nematodes

The phylum **Nematoda**, or roundworms, includes more than 28,000 species with an estimated 16,000 parasitic species. The name Nematoda is derived from the Greek word "nemos," which means "thread." Nematodes are present in all habitats and are extremely common, although they are usually not visible (**Figure 25.25**).

(a)



(b)

Figure 25.25 (a) An scanning electron micrograph of the nematode *Heterodera glycines* and (b) a schematic representation of the anatomy of a nematode are shown. (credit a: modification of work by USDA, ARS; scale-bar data from Matt Russell)

Most nematodes look similar to each other: slender tubes, tapered at each end (**Figure 25.25**). Nematodes are pseudocoelomates and have a **complete digestive system** with a distinct mouth and anus.

The nematode body is encased in a cuticle, a flexible but tough exoskeleton, or external skeleton, which offers protection and support. The cuticle contains a carbohydrate-protein polymer called chitin. The cuticle also lines the pharynx and rectum. Although the exoskeleton provides protection, it restricts growth, and therefore must be continually shed and replaced as the animal increases in size.

A nematode's mouth opens at the anterior end with three or six lips and, in some species, teeth in the form of cuticular extensions. There may also be a sharp stylet that can protrude from the mouth to stab prey or pierce plant or animal cells. The mouth leads to a muscular pharynx and intestine, leading to the rectum and anal opening at the posterior end.

Physiological Processes of Nematodes

In nematodes, the excretory system is not specialized. Nitrogenous wastes are removed by diffusion. In marine nematodes, regulation of water and salt is achieved by specialized glands that remove unwanted ions while maintaining internal body fluid concentrations.

Most nematodes have four nerve cords that run along the length of the body on the top, bottom, and sides. The nerve cords fuse in a ring around the pharynx, to form a head ganglion or "brain" of the worm, as well as at the posterior end to form the tail ganglion. Beneath the epidermis lies a layer of longitudinal muscles that permits only side-to-side, wave-like undulation of the body.

Nematodes employ a diversity of sexual reproductive strategies depending on the species; they may be monoecious, dioecious (separate sexes), or may reproduce asexually by parthenogenesis. *Caenorhabditis elegans* is nearly unique among animals in having both self-fertilizing hermaphrodites and a male sex that can mate with the hermaphrodite.

Arthropoda

The name of this phylum, " **Arthropoda**" means "jointed legs," which aptly describes each of the enormous number of species belonging to this phylum. An estimated 85 percent of known species belong to this phylum, with many more still undiscovered or undescribed. Insects form the largest single group within the phylum. The principal characteristics of all the animals in this phylum are functional segmentation of the body and the presence of jointed appendages (**Figure 25.26**). Arthropods also have an exoskeleton made principally of chitin. Arthropods are true coelomate animals and exhibit protostomic development.

Figure 25.26 Trilobites, like the one in this fossil, are an extinct group of arthropods. (credit: Kevin Walsh)

Physiological Processes of Arthropods

A unique feature of arthropods is the presence of a segmented body with fusion of certain sets of segments to give rise to functional segments. Fused segments may form a head, thorax, and abdomen, or a cephalothorax and abdomen, or a head and trunk. The coelom takes the form of a hemocoel (or blood cavity). The open circulatory system, in which blood bathes the internal organs rather than circulating in vessels, is powered by a two-chambered heart. Respiratory systems vary, depending on the group of arthropod: Insects and myriapods use a series of tubes (tracheae) that branch throughout the body, open to the outside through openings called spiracles, and perform gas exchange directly between the cells and air in the tracheae. Aquatic crustaceans use gills, arachnids employ "book lungs," and aquatic chelicerates use "book gills." The book lungs of arachnids are internal stacks of alternating air pockets and hemocoel tissue shaped like the pages of a book. The book gills of crustaceans are external structures similar to book lungs with stacks of leaf-like structures that exchange gases with the surrounding water (**Figure 25.27**).

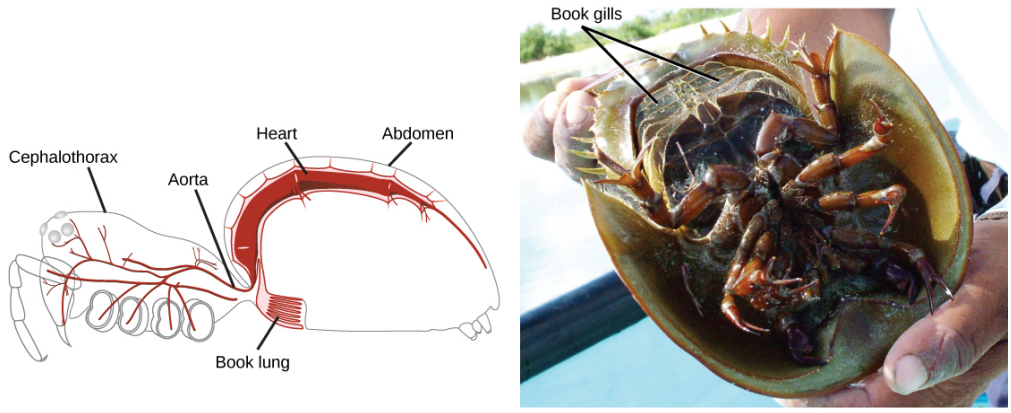


Figure 25.27 The book lungs of (a) arachnids are made up of alternating air pockets and hemocoel tissue shaped like a stack of books. The book gills of (b) crustaceans are similar to book lungs but are external so that gas exchange can occur with the surrounding water. (credit a: modification of work by Ryan Wilson based on original work by John Henry Comstock; credit b: modification of work by Angel Schatz)

Arthropod Diversity

Phylum Arthropoda includes animals that have been successful in colonizing terrestrial, aquatic, and aerial habitats. The phylum is further classified into five subphyla: Trilobitomorpha (trilobites), Hexapoda (insects and relatives), Myriapoda (millipedes, centipedes, and relatives), Crustacea (crabs, lobsters, crayfish, isopods, barnacles, and some zooplankton), and Chelicerata (horseshoe crabs, arachnids, scorpions, and daddy longlegs). Trilobites are an extinct group of arthropods found from the Cambrian period (540–490 million years ago) until they became extinct in the Permian (300–251 million years ago) that are probably most closely related to the Chelicerata. The 17,000 described species have been identified from fossils (**Figure 25.26**).

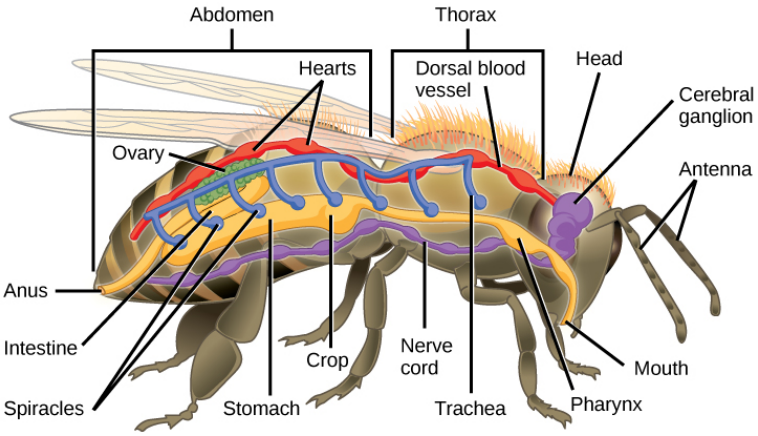


Figure 25.28 The basic anatomy of a representative arthropod, in this case an insect, or hexapod. Note that insects have a developed digestive system (yellow), a respiratory system (blue), a circulatory system (red), and a nervous system (purple).

25.5 | Mollusks and Annelids

Introduction

> “Clams are very conservative. They voted against having heads in the Ordovician Period and have stuck to it ever since.”

Will Cuppy, American humorist, in *How to Attract the Wombat*, 1949

The mollusks are a diverse group (85,000 described species) of mostly marine species. They have a variety of forms, ranging from large predatory squid and octopus, some of which show a high degree of intelligence, to small grazing forms with elaborately sculpted and colored shells. The annelids traditionally include the oligochaetes, which include the earthworms and leeches, the polychaetes, which are a marine group, and two other smaller classes.

Phylum Mollusca

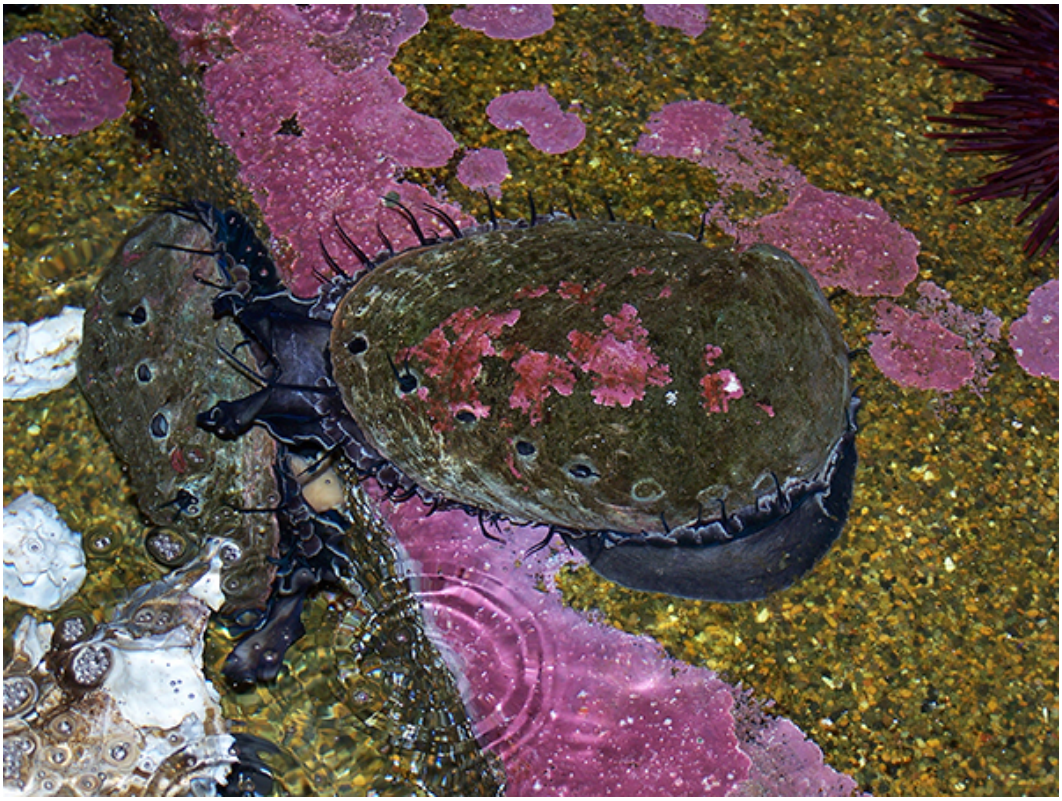


Figure 25.29 An Edible Mollusk Some mollusks are edible, like these Black Abalone (*Haliotis cracherodii*), cruising in a tide pool of the coast of California. This gastropod mollusk shows several of the features of many members of this phylum, including a sturdy shell and a muscular foot. However, due to overharvesting of this species, its populations are declining and it can no longer be legally taken. "Abalone OCA" by Little Mountain 5 - Own work. Licensed under Creative Commons Attribution-Share Alike 3.0 via Wikimedia Commons.

Mollusca is the predominant phylum in marine environments; it is estimated that 23 percent of all known marine species belong to this phylum. It is the second most diverse phylum of animals with over 75,000 described species. The name “mollusca” signifies a soft body, as the earliest descriptions of mollusks came from observations of unshelled, soft-bodied cuttlefish (squid relatives). Although mollusk body forms vary, they share key characteristics, such as a ventral, muscular foot that is typically used for locomotion; the visceral mass, which contains most of the internal organs of the animal; and a dorsal mantle, which is a flap of tissue over the visceral mass that creates a space called the mantle cavity. The mantle may or may not secrete a shell of calcium carbonate. In addition, many mollusks have a scraping structure at the mouth, called a radula (**Figure 25.30**).

The muscular foot varies in shape and function, depending on the type of mollusk. It is a retractable as well as extendable organ, used for locomotion and anchorage. Mollusks are eucoelomates, but the coelomic cavity is restricted to a cavity around the heart in adult animals. The mantle cavity, formed inside the mantle, develops independently of the coelomic cavity. It is a multi-purpose space, housing the gills, the anus, organs for sensing food particles in the water, and an outlet for gametes. Most mollusks have an open circulatory system with a heart that circulates the hemolymph in open spaces around the organs. The octopuses and squid are an exception to this and have a closed circulatory system with two hearts that move blood through the gills and a third, systemic heart that pumps blood through the rest of the body.

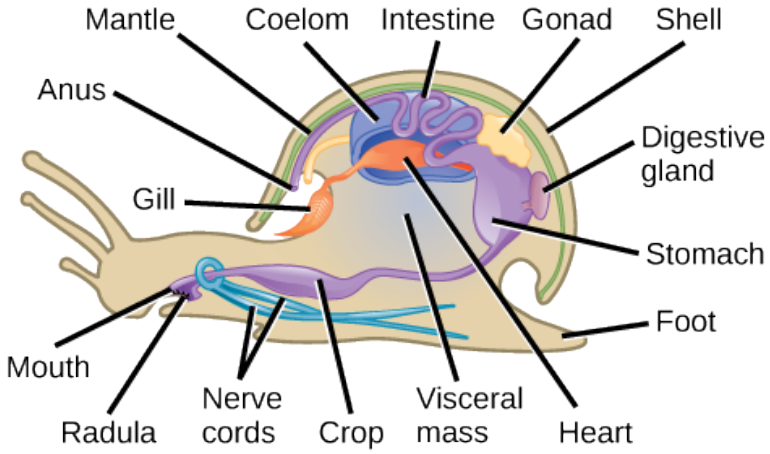


Figure 25.30 There are many species and variations of mollusks; the gastropod mollusk anatomy is shown here, which shares many characteristics common with other groups.

Annelida

Phylum **Annelida** are segmented worms found in marine, terrestrial, and freshwater habitats, but the presence of water or humidity is a critical factor for their survival in terrestrial habitats. The name of the phylum is derived from the Latin word *annellus*, which means a small ring. Approximately 16,500 species have been described. The phylum includes earthworms, polychaete worms, and leeches. Like mollusks, annelids exhibit protostomic development.

Annelids are bilaterally symmetrical and have a worm-like appearance. Their particular segmented body plan results in repetition of internal and external features in each body segment. This type of body plan is called metamerism. The evolutionary benefit of such a body plan is thought to be the capacity it allows for the evolution of independent modifications in different segments that perform different functions. The overall body can then be divided into head, body, and tail.

Physiological Processes of Annelida

The skin of annelids is protected by a cuticle that is thinner than the cuticle of nematodes and arthropods, and it does not need to be molted for growth. Chitinous hairlike extensions, anchored in the skin and projecting from the cuticle, called chaetae, are present in every segment in most groups. The chaetae are a defining character of annelids. Beneath the cuticle there are two layers of muscle, one running around its circumference (circular) and one running the length of the worm (longitudinal). Annelids have a true coelom in which organs are distributed and bathed in coelomic fluid. Annelids possess a well-developed complete digestive system with specialized organs: mouth, muscular pharynx, esophagus, and crop. A cross-sectional view of a body segment of an earthworm is shown in **Figure 25.31**; each segment is limited by a membrane that divides the body cavity into compartments.

Annelids have a closed circulatory system with muscular pumping “hearts” in the anterior segments, dorsal and ventral blood vessels that run the length of the body with connections in each segment, and capillaries that service individual tissues. Gas exchange occurs across the moist body surface. Excretion is carried out by pairs of primitive “kidneys” called metanephridia that consist of a convoluted tubule and an open, ciliated funnel present in every segment. Annelids have a well-developed nervous system with two ventral nerve cords and a nerve ring of fused ganglia present around the pharynx.

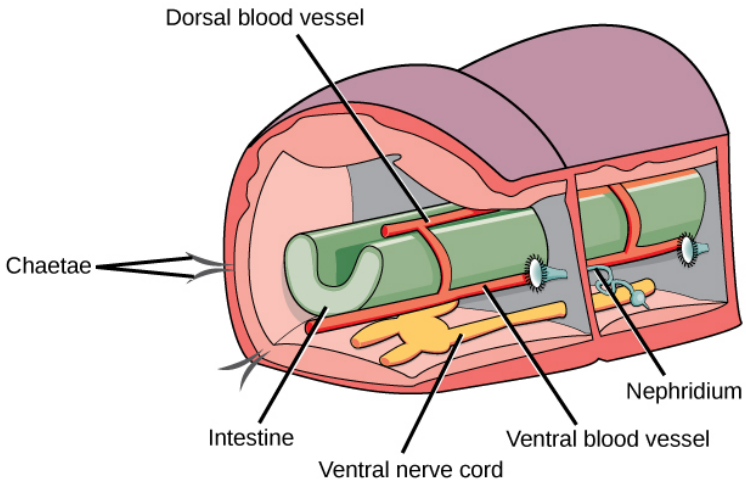


Figure 25.31 In this schematic showing the basic anatomy of annelids, the digestive system is indicated in green, the nervous system is indicated in yellow, and the circulatory system is indicated in red.

Annelids may be either monoecious (hermaphroditic, capable of producing both male and female gametes), such as earthworms and leeches) or dioecious (indivduals are either male or female), as is the case for polychaetes.

25.6 | Echinoderms and Chordates

Introduction

> "Here's a little lesson in deuterostome taxonomy for everyone out there. These are animals in which the first embryonic opening become the anus, and the second opening becomes the mouth (the name literally means mouth second). This is in contrast the majority of animals, which form their mouth first."

RPM, author of the Evolgen blog, 2008

The phyla Echinodermata and Chordata (which includes the vertebrates) and two smaller phyla are the members of the Deuterostomes. Deuterostomes share similar patterns of early development, as noted above, which distinguish them from most of the other animals.

Echinoderms

Echinodermata are named for their spiny skin (from the Greek "echinos" meaning "spiny" and "dermos" meaning "skin"). The phylum includes about 7,000[1] described living species, such as sea stars, sea cucumbers, sea urchins, sand dollars, and brittle stars. **Echinodermata** are exclusively marine.

Although the early larval stages of all echinoderms have bilateral symmetry, adult echinoderms exhibit pentaradial symmetry and have an endoskeleton made of ossicles (**Figure 25.32**). The endoskeleton is developed by epidermal cells, which may also possess pigment cells, giving vivid colors to these animals, as well as cells laden with toxins. These animals have a true coelom, a portion of which is modified into a unique circulatory system called a water vascular system. An interesting feature of these animals is their power to regenerate, even when over 75 percent of their body mass is lost.

1. "Number of Living Species in Australia and the World," A.D. Chapman, Australia Biodiversity Information Services, last modified August 26, 2010, http://www.environment.gov.au/biodiversity/abrs/publications/other/species-numbers/2009/03-exec-summary.html.

Physiological Processes of Echinoderms

Echinoderms have a unique system for gas exchange, nutrient circulation, and locomotion called the water vascular system. The system consists of a central ring canal and radial canals extending along each arm. Water circulates through these structures allowing for gas, nutrient, and waste exchange. A structure on top of the body, called the madreporite, regulates the amount of water in the water vascular system. “Tube feet,” which protrude through openings in the endoskeleton, may be expanded or contracted using the hydrostatic pressure in the system. The system allows for slow movement, but a great deal of power, as witnessed when the tube feet latch on to opposite halves of a bivalve mollusk, like a clam, and slowly, but surely pull the shells apart, exposing the flesh within.

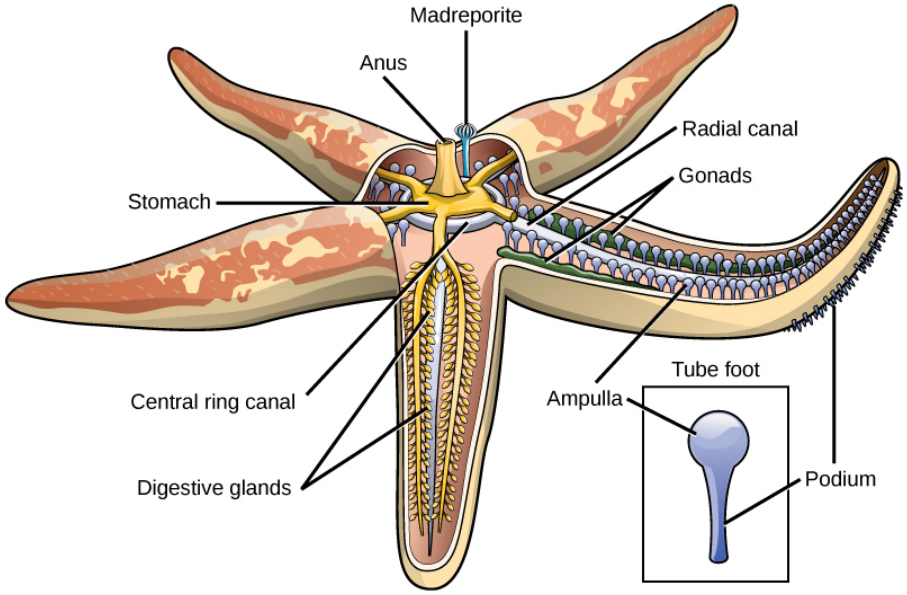


Figure 25.32 This diagram shows the anatomy of a sea star.

The echinoderm nervous system has a nerve ring at the center and five radial nerves extending outward along the arms. There is no centralized nervous control. Echinoderms have separate sexes and release their gametes into the water where fertilization takes place. Echinoderms may also reproduce asexually through regeneration from body parts.

Echinoderm Diversity

This phylum is divided into five classes: Asteroidea (sea stars), Ophiuroidea (brittle stars), Echinoidea (sea urchins and sand dollars), Crinoidea (sea lilies or feather stars), and Holothuroidea (sea cucumbers) (**Figure 25.33**).

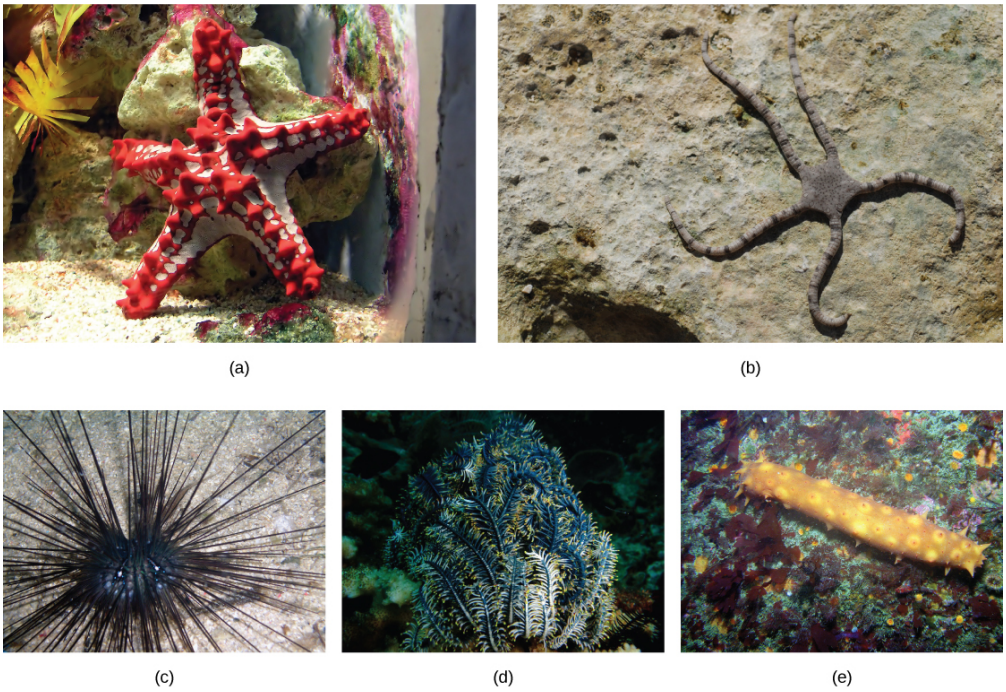


Figure 25.33 Different members of Echinodermata include the (a) sea star in class Asteroidea, (b) the brittle star in class Ophiuroidea, (c) the sea urchins of class Echinoidea, (d) the sea lilies belonging to class Crinoidea, and (d) sea cucumbers representing class Holothuroidea. (credit a: modification of work by Adrian Pingstone; credit b: modification of work by Joshua Ganderson; credit c: modification of work by Samuel Chow; credit d: modification of work by Sarah Depper; credit e: modification of work by Ed Bierman)

Chordates

The majority of species in the phylum Chordata are found in the subphylum Vertebrata, which include many species with which we are familiar. The vertebrates contain more than 60,000 described species, divided into major groupings of the lampreys, fishes, amphibians, reptiles, birds, and mammals.

Animals in the phylum **Chordata** share four key features that appear at some stage of their development: a notochord, a dorsal hollow nerve cord, pharyngeal slits, and a post-anal tail (**Figure 25.34**). In certain groups, some of these traits are present only during embryonic development.

The chordates are named for the **notochord**, which is a flexible, rod-shaped structure that is found in the embryonic stage of all chordates and in the adult stage of some chordate species. It is located between the digestive tube and the nerve cord, and provides skeletal support through the length of the body. In some chordates, the notochord acts as the primary axial support of the body throughout the animal's lifetime. In vertebrates, the notochord is present during embryonic development, at which time it induces the development of the neural tube and serves as a support for the developing embryonic body. The notochord, however, is not found in the postnatal stage of vertebrates; at this point, it has been replaced by the vertebral column (the spine).

The **dorsal hollow nerve cord** is derived from ectoderm that sinks below the surface of the skin and rolls into a hollow tube during development. In chordates, it is located dorsally to the notochord. In contrast, other animal phyla possess solid nerve cords that are located either ventrally or laterally. The nerve cord found in most chordate embryos develops into the brain and spinal cord, which compose the central nervous system.

Pharyngeal slits are openings in the pharynx, the region just posterior to the mouth, that extend to the outside environment. In organisms that live in aquatic environments, pharyngeal slits allow for the exit of water that enters the mouth during feeding. Some invertebrate chordates use the pharyngeal slits to filter food from the water that enters the mouth. In fishes, the pharyngeal slits are modified into gill supports, and in jawed fishes, jaw supports. In tetrapods, the slits are further modified into components of the ear and tonsils, since there is no longer any need for gill supports in these air-breathing animals. Tetrapod means "four-footed," and this group includes amphibians, reptiles, birds, and mammals. (Birds are considered tetrapods because they evolved from tetrapod ancestors.)

The **post-anal tail** is a posterior elongation of the body extending beyond the anus. The tail contains skeletal elements and muscles, which provide a source of locomotion in aquatic species, such as fishes. In some terrestrial vertebrates, the tail

may also function in balance, locomotion, courting, and signaling when danger is near. In many species, the tail is absent or reduced; for example, in apes, including humans, it is present in the embryo, but reduced in size and nonfunctional in adults.

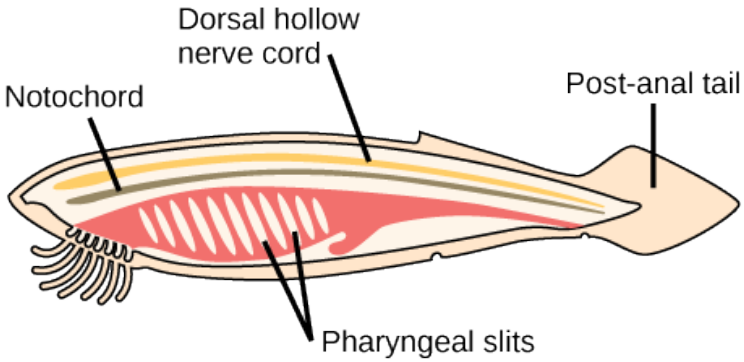


Figure 25.34 In chordates, four common features appear at some point in development: a notochord, a dorsal hollow nerve cord, pharyngeal slits, and a post-anal tail. The anatomy of a cephalochordate shown here illustrates all of these features.

Invertebrate Chordates

In addition to the vertebrates, the phylum Chordata contains two clades of invertebrates: Urochordata (tunicates) and Cephalochordata (lancelets). Members of these groups possess the four distinctive features of chordates at some point during their development.

The tunicates (**Figure 25.35**) are also called sea squirts. The name tunicate derives from the cellulose-like carbohydrate material, called the tunic, which covers the outer body. Although tunicates are classified as chordates, the adult forms are much modified in body plan and do not have a notochord, a dorsal hollow nerve cord, or a post-anal tail, although they do have pharyngeal slits. The larval form possesses all four structures. Most tunicates are hermaphrodites. Tunicate larvae hatch from eggs inside the adult tunicate's body. After hatching, a tunicate larva swims for a few days until it finds a suitable surface on which it can attach, usually in a dark or shaded location. It then attaches by the head to the substrate and undergoes metamorphosis into the adult form, at which point the notochord, nerve cord, and tail disappear.

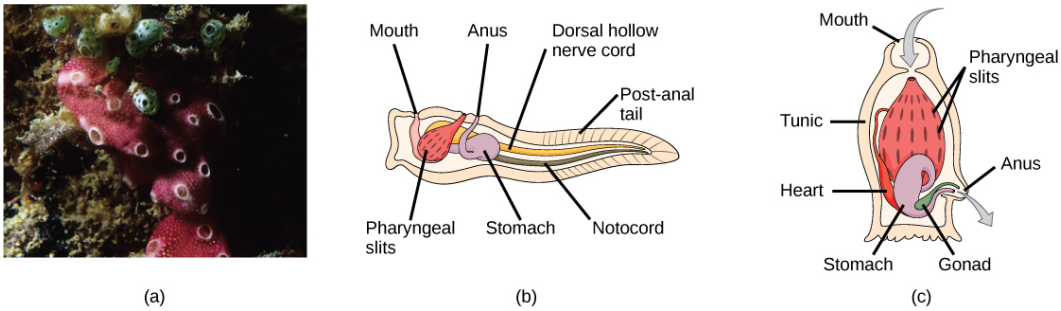


Figure 25.35 (a) This photograph shows a colony of the tunicate *Botrylloides violaceus*. In the (b) larval stage, the tunicate can swim freely until it attaches to a substrate to become (c) an adult. (credit a: modification of work by Dr. Dwayne Meadows, NOAA/NMFS/OPR)

Lancelets possess a notochord, dorsal hollow nerve cord, pharyngeal slits, and a post-anal tail in the adult stage (**Figure 25.36**). The notochord extends into the head, which gives the subphylum its name (Cephalochordata). Extinct fossils of this subphylum date to the middle of the Cambrian period (540–488 mya).The living forms, the lancelets, are named for their blade-like shape. Lancelets are only a few centimeters long and are usually found buried in sand at the bottom of warm temperate and tropical seas. Like tunicates, they are suspension feeders.

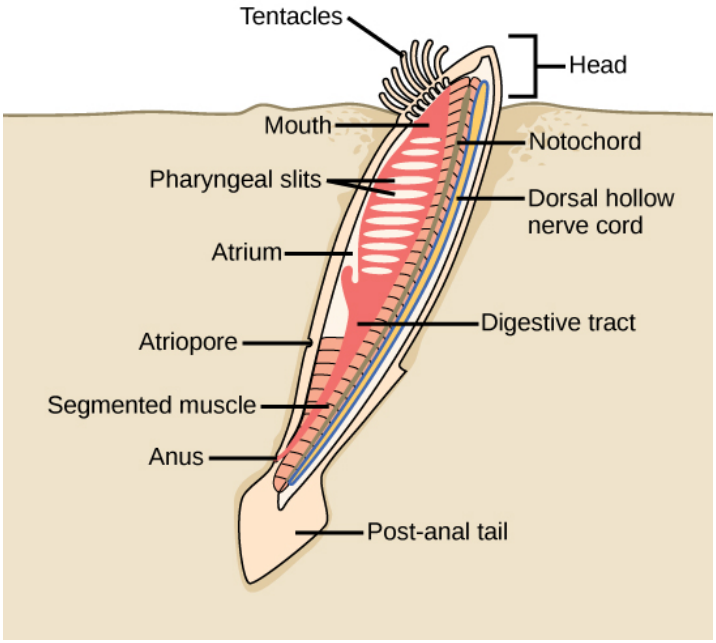


Figure 25.36 Adult lancelets retain the four key features of chordates: a notochord, a dorsal hollow nerve cord, pharyngeal slits, and a post-anal tail.

25.7 | Vertebrates

Introduction

> "The development of the Vertebrate proceeds from an axis upward, in two layers, which coalesce at the edges, and also downward, in two layers, which likewise coalesce at the edges. Thus two main tubes are formed, one above the other. During the formation of these, the embryo separates into strata, so that the two main tubes are composed of subordinate tubes which enclose each other as fundamental organs, and are capable of developing into all the organs."

Karl Ernst von Baer, German embryologist, 1828

Vertebrates are the largest and probably the most recognizable organisms of the animal kingdom (**Figure 25.37**). More than 62,000 vertebrate species have been identified. The vertebrate species now living represent only a small portion of the vertebrates that have existed. The best-known extinct vertebrates are the dinosaurs, a unique group of reptiles, reaching sizes not seen before or since in terrestrial animals.

(a) (b) (c)

Figure 25.37 Examples of critically endangered vertebrate species include (a) the Siberian tiger *(Panthera tigris altaica*), (b) the Panamanian golden frog (*Atelopus zeteki*), and (c) the Philippine eagle (*Pithecophaga jefferyi*). (credit a: modification of work by Dave Pape; credit b: modification of work by Brian Gratwicke; credit c: modification of work by "cuatrok77"/Flickr)

Fishes

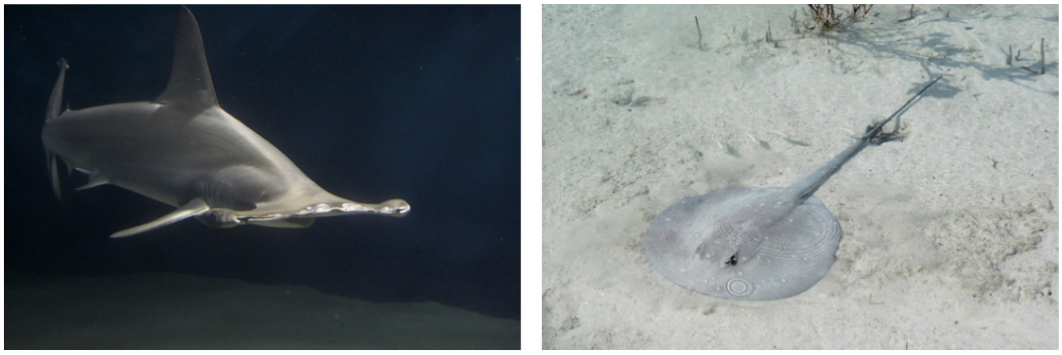
Modern fishes include an estimated 31,000 species. Fishes were the earliest vertebrates, and jawless fishes were the earliest of these. Jawless fishes—the present day hagfishes and lampreys—have a distinct cranium and complex sense organs including eyes, distinguishing them from the invertebrate chordates. The jawed fishes evolved later and are extraordinarily diverse today. These include the cartilaginous fishes (e.g. rays and sharks), with a skeleton made of cartilage, and the bony fishes, which have a bony skeleton. All fishes are active feeders, rather than sessile, suspension feeders.

Jawless fishes

(a) (b)

Figure 25.38 Jawless fishes(a) Jawless fishes include 67 species of hagfishes. These Pacific hagfishes are scavengers that live on the ocean floor. (b) Lampreys are another type of jawless fish. These parasitic sea lampreys attach to their lake trout host by suction and use their rough tongues to rasp away flesh in order to feed on the trout's blood. (credit a: modification of work by Linda Snook, NOAA/CBNMS; credit b: modification of work by USGS)

Jawed fishes



(a) (b)

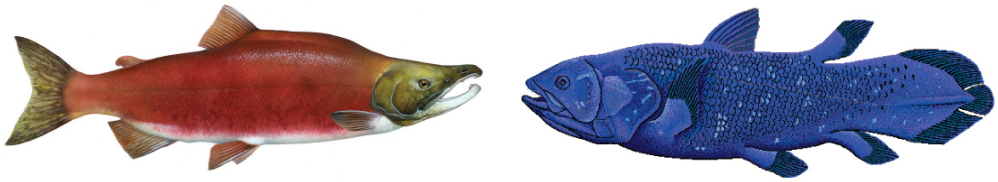
Figure 25.39 Cartilaginous fishes The jawed fishes include the clade Chondrichthyes, fishes with skeletons made of cartilage, and these include the sharks and rays(a) This hammerhead shark is an example of a predatory cartilaginous fish. (b) This stingray blends into the sandy bottom of the ocean floor when it is feeding or awaiting prey. (credit a: modification of work by Masashi Sugawara; credit b: modification of work by "Sailn1"/Flickr)

Sharks reproduce sexually and eggs are fertilized internally. Most species are ovoviviparous, that is, the fertilized egg is retained in the oviduct of the mother's body, and the embryo is nourished by the egg yolk. The eggs hatch in the uterus and young are born alive and fully functional. Some species of sharks are oviparous: They lay eggs that hatch outside of the mother's body. Embryos are protected by a shark egg case or "mermaid's purse" that has the consistency of leather. The shark egg case has tentacles that snag in seaweed and give the newborn shark cover. A few species of sharks are viviparous, that is, the young develop within the mother's body, and she gives live birth.

Rays and skates include more than 500 species and are closely related to sharks. They can be distinguished from sharks by their flattened bodies, pectoral fins that are enlarged and fused to the head, and gill slits on their ventral surface (**Figure 25.39b**). Like sharks, rays and skates have a cartilaginous skeleton. Most species are marine and live on the sea floor, with nearly a worldwide distribution.

Bony Fishes

Members of the clade Osteichthyes, or bony fishes, are characterized by a bony skeleton. The vast majority of present-day fishes belong to this group, which consists of approximately 30,000 species, making it the largest class of vertebrates in existence today.



(a) (b)

Figure 25.40 The (a) sockeye salmon and (b) coelacanth are both bony fishes of the Osteichthyes clade. The coelacanth, sometimes called a lobe-finned fish, was thought to have gone extinct in the Late Cretaceous period 100 million years ago until one was discovered in 1938 between Africa and Madagascar. (credit a: modification of work by Timothy Knepp, USFWS; credit b: modification of work by Robbie Cada)

Amphibians

Amphibians are vertebrate tetrapods. Amphibia includes the familiar frogs, toads, and salamanders, as well as the caecilians, limbless amphibians that superficially resemble worms and snakes. The term amphibian means "dual life," which is a reference to the metamorphosis that many frogs undergo from a tadpole to an adult and the mixture of aquatic and terrestrial environments in their life cycle. Amphibians evolved in the Devonian period and were the earliest terrestrial tetrapods.

Amphibian Diversity

Amphibia comprise an estimated 6,500 extant species that inhabit tropical and temperate regions around the world. Amphibians can be divided into three clades: Urodela ("tailed-ones"), the salamanders and newts; Anura ("tail-less ones"), the frogs and toads; and Apoda ("legless ones"), the caecilians.

(a) (b)

Figure 25.41 Frogs and Salamanders There are about 500 species of salamanders and 5,000 species of frogs worldwide. (a) Salamanders are terrestrial animals, but most are commonly found only near water. (b) The Australian green tree frog is a nocturnal predator that lives in the canopies of trees near a water source. (credit a: modification of work by Valentina Storti; credit b: modification of work by Evan Pickett)

Caecilians comprise an estimated 185 species. They lack external limbs and resemble giant earthworms. They inhabit soil and are found primarily in the tropics of South America, Africa, and southern Asia where they are adapted for a soil-burrowing lifestyle and are nearly blind. Unlike most of the other amphibians that breed in or near water, reproduction in a drier soil habitat means that caecilians must utilize internal fertilization, and most species give birth to live young (**Figure 25.42**).

Figure 25.42 Caecilians lack external limbs and are well adapted for a soil-burrowing lifestyle. (credit: modification of work by "cliff1066"/Flickr)

Reptiles and Birds

The amniotes—reptiles, birds, and mammals—are distinguished from amphibians by their terrestrially adapted (shelled) egg and an embryo protected by amniotic membranes. The evolution of amniotic membranes meant that the embryos of amniotes could develop within an aquatic environment inside the egg. This led to less dependence on a water environment for development and allowed the amniotes to invade drier areas. This was a significant evolutionary change that distinguished them from amphibians, which were restricted to moist environments due to their shell-less eggs. Although the shells of various amniotic species vary significantly, they all allow retention of water. The membranes of the amniotic egg also allowed gas exchange and sequestering of wastes within the enclosure of an eggshell. The shells of bird eggs are composed of calcium carbonate and are hard and brittle, but possess pores for gas and water exchange. The shells of reptile eggs are more leathery and pliable. Most mammals do not lay eggs; however, even with internal gestation, amniotic membranes are still present.

In the past, the most common division of amniotes has been into classes Mammalia, Reptilia, and Aves. Birds are descended, however, from dinosaurs, so this classical scheme results in groups that are not true clades. We will discuss birds as a group distinct from reptiles with the understanding that this does not reflect evolutionary history.

Reptiles

Reptiles are tetrapods. Limbless reptiles—snakes—may have vestigial limbs and, like caecilians, are classified as tetrapods because they are descended from four-limbed ancestors. Reptiles lay shelled eggs on land. Even aquatic reptiles, like sea turtles, return to the land to lay eggs. They usually reproduce sexually with internal fertilization. Some species display ovoviviparity, with the eggs remaining in the mother's body until they are ready to hatch. Other species are viviparous, with the offspring born alive.

One of the key adaptations that permitted reptiles to live on land was the development of their scaly skin, containing the protein keratin and waxy lipids, which prevented water loss from the skin. This occlusive skin means that reptiles cannot use their skin for respiration, like amphibians, and thus all must breathe with lungs. In addition, reptiles conserve valuable body water by excreting nitrogen in the form of uric acid paste. These characteristics, along with the shelled, amniotic egg, were the major reasons why reptiles became so successful in colonizing a variety of terrestrial habitats far from water.

Reptiles are ectotherms, that is, animals whose main source of body heat comes from the environment. Behavioral maneuvers, like basking to heat themselves, or seeking shade or burrows to cool off, help them regulate their body temperature,

Class Reptilia includes diverse species classified into four living clades. These are the Crocodilia, Sphenodontia, Squamata, and Testudines.

The Crocodilia ("small lizard") arose approximately 84 million years ago, and living species include alligators, crocodiles, and caimans. Crocodilians (**Figure 25.43a**) live throughout the tropics of Africa, South America, the southeastern United States, Asia, and Australia. They are found in freshwater habitats, such as rivers and lakes, and spend most of their time in water. Some species are able to move on land due to their semi-erect posture.

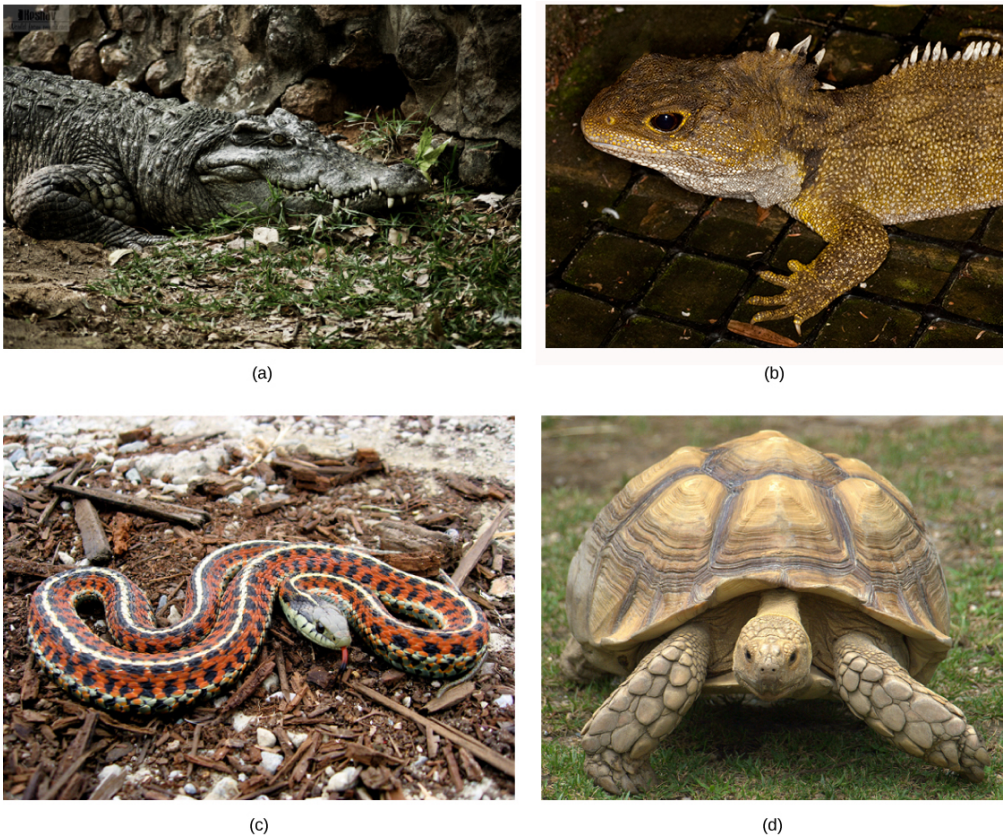


Figure 25.43 (a) Crocodilians, such as this Siamese crocodile, provide parental care for their offspring. (b) This Tuatara (*Sphenodon punctatus* is one of only two species in the Sphenodontia, and is found only in New Zealand. (c) The garter snake belongs to the genus *Thamnophis*, the most widely distributed reptile genus in North America. (d) The African spurred tortoise lives at the southern edge of the Sahara Desert. It is the third largest tortoise in the world. (credit a: modification of work by Keshav Mukund Kandhadai; credit b: courtesy of David A. Rintoul; credit c: modification of work by Steve Jurvetson; credit d: modification of work by Jim Bowen)

The Sphenodontia ("wedge tooth") arose in the Mesozoic Era and includes only one living genus, the *Tuatara*, with two species that are found in New Zealand (**Figure 25.43b**). There are many fossil species extending back to the Triassic period (250–200 million years ago). Although the tuataras resemble lizards, they are anatomically distinct and share characteristics that are found in birds and turtles.

Squamata ("scaly") arose in the late Permian; living species include lizards and snakes, which are the largest extant clade of reptiles. Lizards differ from snakes by having four limbs, eyelids, and external ears, which are lacking in snakes. Lizard species range in size from chameleons and geckos that are a few centimeters in length to the Komodo dragon, which is about 3 meters in length.

Snakes are thought to have descended from either burrowing lizards or aquatic lizards over 100 million years ago (**Figure 25.43c**). Snakes comprise about 3,000 species and are found on every continent except Antarctica. They range in size from 10 centimeter-long thread snakes to 7.5 meter-long pythons and anacondas. All snakes are carnivorous and eat small animals, birds, eggs, fish, and insects.

Turtles are members of the clade Testudines ("having a shell") (**Figure 25.43d**). Turtles are characterized by a bony or cartilaginous shell, made up of the carapace on the back and the plastron on the ventral surface, which develops from the ribs. Turtles arose approximately 200 million years ago, predating crocodiles, lizards, and snakes. Turtles lay eggs on land, although many species live in or near water. Turtles range in size from the speckled padloper tortoise at 8 centimeters (3.1 inches) to the leatherback sea turtle at 200 centimeters (over 6 feet). The term "turtle" is sometimes used to describe only those species of Testudines that live in the sea, with the terms "tortoise" and "terrapin" used to refer to species that live on land and in fresh water, respectively.

Birds

Fossil and genomic data now suggest that birds belong within the reptile clade, but they display a number of unique adaptations that set them apart. Unlike the reptiles, birds are endothermic, meaning they generate their own body heat through metabolic processes. The most distinctive characteristic of birds is their feathers, which are modified reptilian scales. Birds have several different types of feathers that are specialized for specific functions, like contour feathers that streamline the bird's exterior and loosely structured down feathers that insulate (**Figure 25.44a**). Current thinking is that some dinosaurs (the ancestors of modern birds) were also endothermic.

Feathers not only permitted the earliest birds to glide, and ultimately engage in flapping flight, but they insulated the bird's body, assisting the maintenance of endothermy, even in cooler temperatures. Powering a flying animal requires economizing on the amount of weight carried. As body weight increases, the muscle output and energetic cost required for flying increase. Birds have made several modifications to reduce body weight, including hollow or pneumatic bones (**Figure 25.44b**) with air spaces that may be connected to air sacs and cross-linked struts within their bones to provide structural reinforcement. Parts of the vertebral skeleton and braincase are fused to increase its strength while lightening its weight. Most species of bird only possess one ovary rather than two, and no living birds have teeth in their jaw, further reducing body mass.

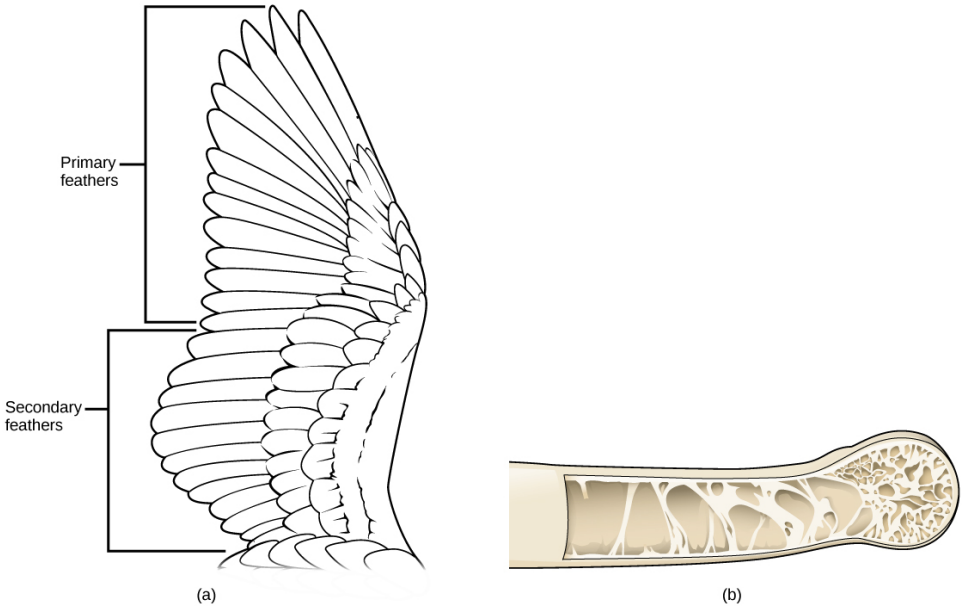


Figure 25.44 (a) Primary feathers are located at the wing tip and provide thrust; secondary feathers are located close to the body and provide lift. (b) Many birds have hollow pneumatic bones, which make flight easier.

Birds possess a system of air sacs branching from their primary airway that divert the path of air so that it passes unidirectionally through the lung, during both inspiration and expiration. Unlike mammalian lungs in which air flows in two directions as it is breathed in and out, air flows continuously through the bird's lung to provide a more efficient system of gas exchange.

Mammals

Mammals are vertebrates that have hair and mammary glands used to provide nutrition for their young. Certain features of the jaw, skeleton, skin, and internal anatomy are also unique to mammals. The presence of hair is one of the key

characteristics of a mammal. Although it is not very extensive in some groups, such as whales, hair has many important functions for mammals. Mammals are endothermic, and hair provides insulation by trapping a layer of air close to the body to retain metabolic heat. Hair also serves as a sensory mechanism through specialized hairs called vibrissae, better known as whiskers. These attach to nerves that transmit touch information, which is particularly useful to nocturnal or burrowing mammals. Hair can also provide protective coloration.

The skeletal system of mammals possesses unique features that differentiate them from other vertebrates. Most mammals have heterodont teeth, meaning they have different types and shapes of teeth that allow them to feed on different kinds of foods. These different types of teeth include the incisors, the canines, premolars, and molars. The first two types are for cutting and tearing, whereas the latter two types are for crushing and grinding. Different groups have different proportions of each type, depending on their diet. Most mammals are also diphyodonts, meaning they have two sets of teeth in their lifetime: deciduous or "baby" teeth, and permanent teeth. In other vertebrates, the teeth can be replaced throughout life.

Modern mammals are divided into three broad groups: monotremes, marsupials, and eutherians (mammals with a placenta). The eutherians and the marsupials collectively are called therian mammals, whereas monotremes are called metatherians.

There are three living species of monotremes: the platypus and two species of echidnas, or spiny anteaters (**Figure 25.45**). The platypus and one species of echidna are found in Australia, whereas the other species of echidna is found in New Guinea. Monotremes are unique among mammals, as they lay leathery eggs, similar to those of reptiles, rather than giving birth to live young. However, the eggs are retained within the mother's reproductive tract until they are almost ready to hatch. Once the young hatch, the female begins to secrete milk from pores in a ridge of mammary tissue along the ventral side of her body. Like other mammals, monotremes are endothermic but regulate body temperatures somewhat lower (90 °F, 32 °C) than placental mammals do (98 °F, 37 °C).

Figure 25.45 The platypus (left), a monotreme, possesses a leathery beak and lays eggs rather than giving birth to live young. An echidna, another monotreme, is shown in the right photo. (credit "echidna": modification of work by Barry Thomas)

Marsupials are found primarily in Australia and nearby islands, although about 100 species of opossums and a few species of two other families are found in the Americas. Australian marsupials number over 230 species and include the kangaroo, koala, bandicoot, and Tasmanian devil (**Figure 25.46**). Most species of marsupials possess a pouch in which the young reside after birth, receiving milk and continuing to develop. Before birth, marsupials have a less complex placental connection, and the young are born much less developed than in placental mammals.

Figure 25.46 The Tasmanian devil is one of several marsupials native to Australia. (credit: Wayne McLean)

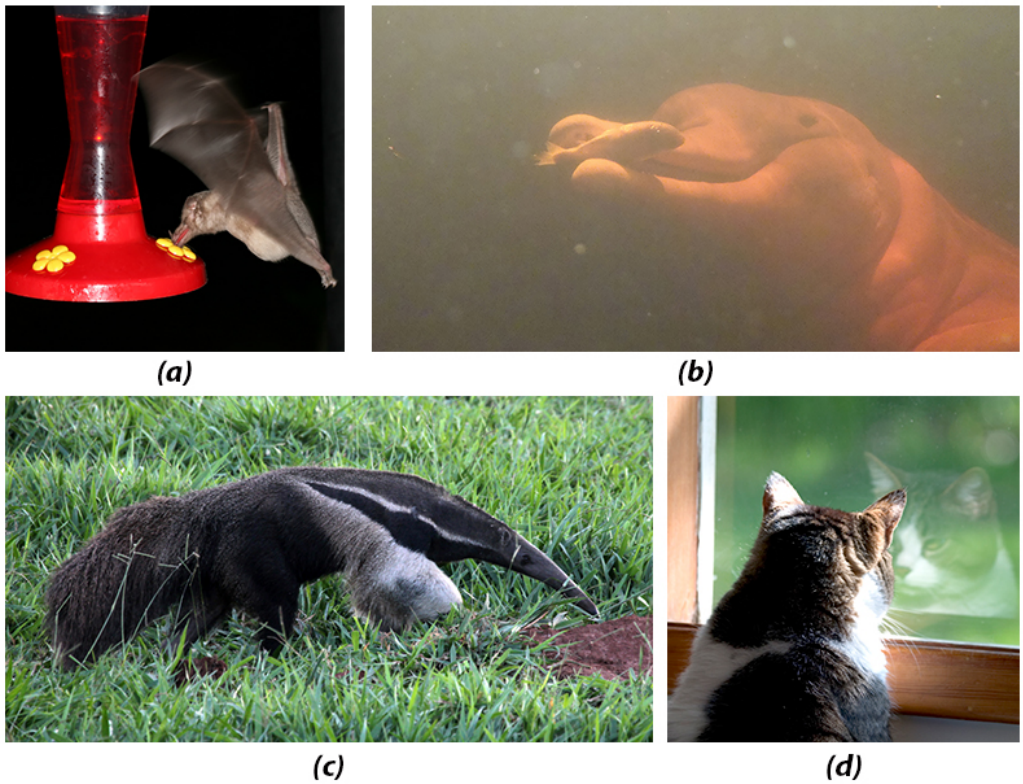


Figure 25.47 Eutherian Diversity Four eutherian mammals: A. A chiropteran, the Mexican Long-tongued Bat (*Choeronycteris mexicana*) feeding on sugar water at a hummingbird feeder. These bats pollinate many varieties of cactuses and agaves, including the blue agave that is used to make tequila. B. Pink River Dolphin (*Inia geoffrensis*), a freshwater cetacean found in the Amazon River and its tributaries. C. A xenarthran, the Giant Anteater or Tamandua (*Tamandua tetradactyla*), hunting for termites in the Pantanal of Brazil. D. A member of the order Carnivora, the Domestic Cat (*Felis catus*). Photo credits - David A. Rintoul

Eutherians are the most widespread of the mammals, occurring throughout the world. There are several groups of eutherians, including Insectivora, the insect eaters; Xenarthra, the toothless anteaters; Rodentia, the rodents; Chiroptera, the bats; Cetacea, the aquatic mammals including whales; Carnivora, carnivorous mammals including dogs, cats, and bears; and Primates, which includes humans. Eutherian mammals are sometimes called placental mammals, because all species have a complex placenta that connects a fetus to the mother, allowing for gas, fluid, waste, and nutrient exchange. While other mammals may possess a less complex placenta or briefly have a placenta, all eutherians have a complex placenta during

gestation.

25.8 | Homeostasis

> "The constant conditions which are maintained in the body might be termed equilibria. That word, however, has come to have a fairly exact meaning as applied to relatively simple physico-chemical states, in closed systems, where known forces are balanced. The coordinated physiological processes which maintain most of the steady states in the organism are so complex, and so peculiar to living beings - involving, as they may, the brain and nerves, the heart, lungs, kidneys and spleen, all working cooperatively - that I have suggested a special designation for these states: **homeostasis**. The word does not imply something set and immobile, a stagnation. It means a condition - a condition which may vary, but which is relatively constant."

Walter Cannon, *The Wisdom of the Body*, 1932, p. 24.

Cannon was an American physiologist who coined and popularized the concept of homeostasis to describe the observations that animals could maintain stable internal body conditions even when the external conditions changed. Animal organs and organ systems constantly adjust to internal and external changes through a process called homeostasis ("steady state"). These changes might be in the level of glucose or calcium in blood or in external temperatures. **Homeostasis** means to maintain dynamic equilibrium in the body. It is dynamic because it is constantly adjusting to the changes that the body's systems encounter. It is equilibrium because body functions are kept within specific ranges. Even an animal that is apparently inactive is maintaining this homeostatic equilibrium.

Homeostatic Process

The goal of homeostasis is the maintenance of equilibrium around a point or value 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 parameter toward the set point. For instance, if the body becomes too warm, adjustments are made to cool the animal. If blood glucose concentration rises after a meal, adjustments are made to lower the blood glucose level, increasing uptake of glucose from blood into various tissues where it can be converted to storage products like glycogen or triglyceride.

Control of Homeostasis

When a change occurs in an animal's environment, an adjustment must be made. The receptor senses the change in the environment, then sends a signal to the control center (in most cases, the brain) which in turn generates a response that is signaled to an effector. The effector is a muscle (that contracts or relaxes) or a gland that secretes. Homeostatsis is maintained by **negative feedback** loops. **Positive feedback** loops actually push the organism further out of homeostasis, but may be necessary for life to occur. Homeostasis is controlled by the nervous and endocrine system of mammals, as described by Cannon in the 1930's.

Negative Feedback Mechanisms

Any homeostatic process that changes the direction of the stimulus is a **negative feedback loop**. It can either cause an increase, or a decrease, in the level of the stimulus that triggered the response, In all cases the response is in the opposite direction of the change in the stimulus. In other words, if a level is too high, the body does something to bring it down, and conversely, if a level is too low, the body does something to make it go up. Hence the term negative feedback. An example is animal maintenance of blood glucose levels, as mentioned above. When an animal has eaten, blood glucose levels rise.

This is sensed by the nervous system. Specialized cells in the pancreas sense this, and the hormone insulin is released by the endocrine system. Insulin causes blood glucose levels to decrease, as would be expected in a negative feedback system, as illustrated in **Figure 25.48**. However, if an animal has not eaten and blood glucose levels decrease, this is sensed in another group of cells in the pancreas, and the hormone glucagon is released, causing glucose levels to increase. This is still a negative feedback loop, which is defined as a situation where a change in one direction is countered by a response in the opposite direction. Negative feedback loops are the predominant mechanism used in homeostasis.

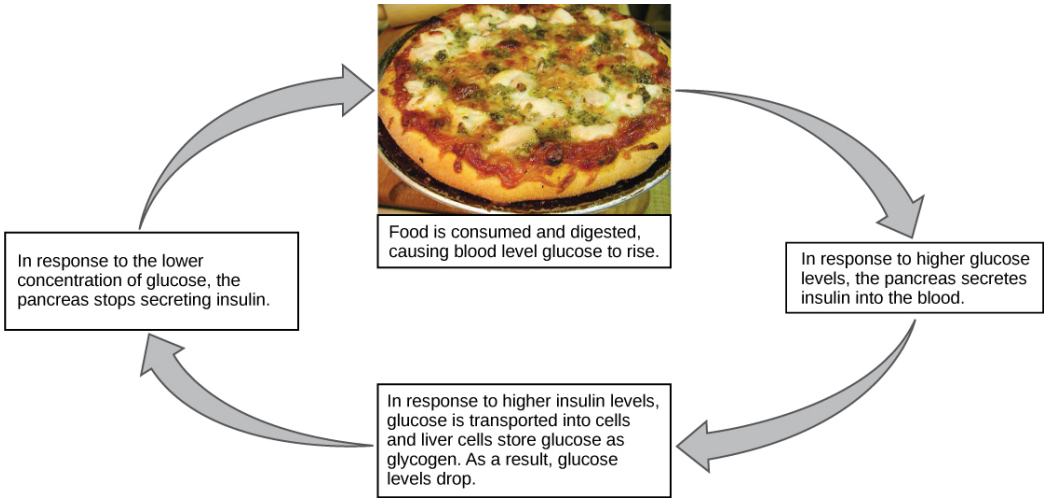


Figure 25.48 Blood sugar levels are controlled by a negative feedback loop. (credit: modification of work by Jon Sullivan)

Positive Feedback Loop

A **positive feedback loop** maintains the direction of the stimulus, possibly accelerating it. Few examples of positive feedback loops exist in animal bodies, but one is found in the cascade of chemical reactions that result in blood clotting, or coagulation. As one clotting factor is activated, it activates the next factor in sequence until a fibrin clot is achieved. The direction is maintained, not changed, so this is positive feedback. Another example of positive feedback is uterine contractions during childbirth, as illustrated in **Figure 25.49**. The pituitary hormone oxytocin stimulates the contraction of the uterus. This produces pain, which is sensed by the nervous system. Instead of lowering the oxytocin and causing the pain to subside, the nervous system causes the pituitary to secrete more oxytocin, stimulating stronger contractions until the contractions are powerful enough to produce childbirth.

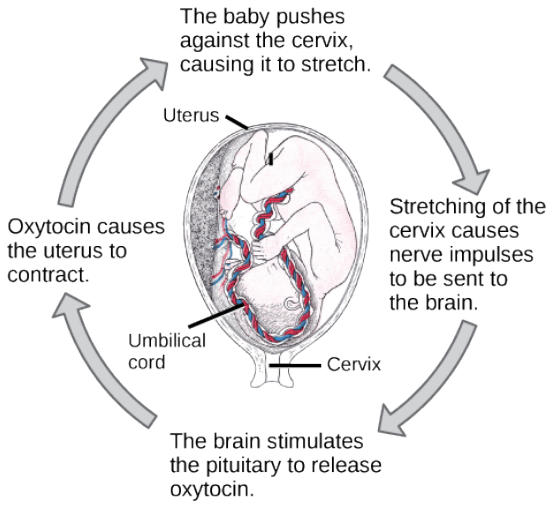


Figure 25.49 The birth of a human infant is the result of positive feedback.

Set Point

It is possible to adjust a system's **set point**, i.e., the level around which the parameter of interest fluctuates.. When this happens, the feedback loop works to maintain the new setting. An example of this is blood pressure: over time, the normal or set point for blood pressure can increase as a result of continued increases in blood pressure. The body no longer recognizes the elevation as abnormal and no attempt is made to return to the lower set point. The result is the maintenance of an

elevated blood pressure that can have harmful effects on the body. Medication can lower blood pressure and lower the set point in the system to a more healthy level.

Changes can be made in a group of body organ systems in order to maintain a set point in another system. This is called acclimatization. This occurs, for instance, when an animal migrates to a higher altitude than it is accustomed to. In order to adjust to the lower oxygen levels at the new altitude, the body increases the number of red blood cells circulating in the blood to ensure adequate oxygen delivery to the tissues. Another example of acclimatization is animals that have seasonal changes in their coats: a heavier coat in the winter ensures adequate heat retention, and a light coat in summer assists in keeping body temperature from rising to harmful levels.

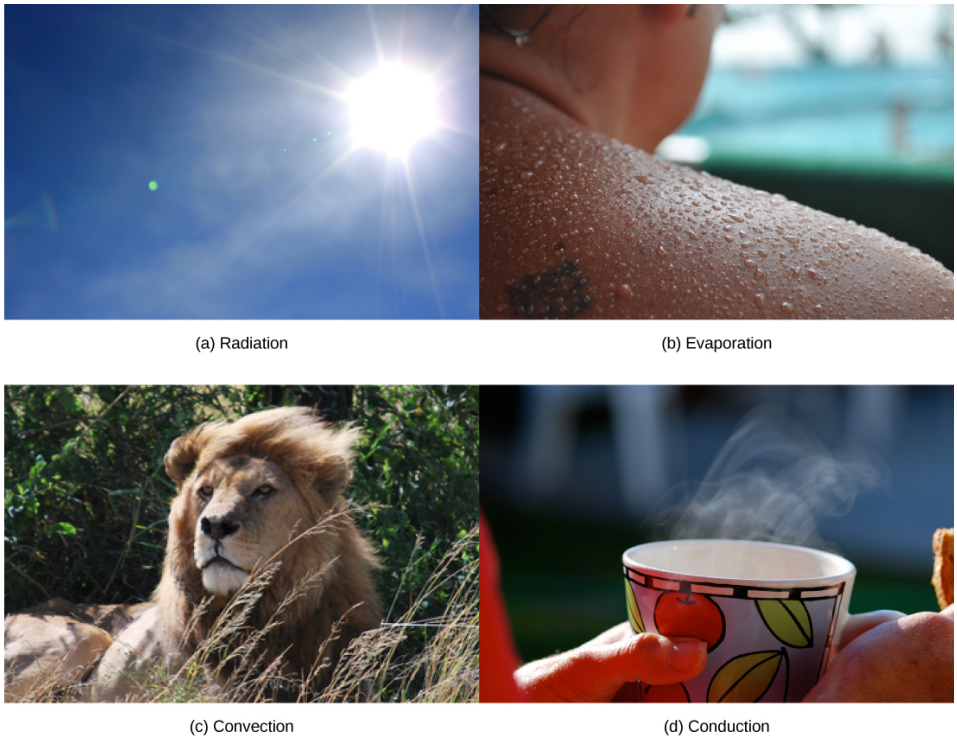
Homeostasis: Thermoregulation

Body temperature affects body activities. Generally, as body temperature rises, enzyme activity rises as well. For every ten degree centigrade rise in temperature, enzyme activity doubles, up to a point. Body proteins, including enzymes, begin to denature and lose their function at even higher temperatures, as you learned in a previous module. Enzyme activity will also decrease by half for every ten degree centigrade drop in temperature, to the point of freezing, with a few exceptions. Some fish can withstand freezing solid and return to normal with thawing, and one mammal (the Arctic ground squirrel *Urocitellus parryii*) can lower its body temperature to -3° C during its winter hibernation.

Endotherms and Ectotherms

Animals can be divided into two groups: some maintain a constant body temperature in the face of differing environmental temperatures, while others have a body temperature that is the same as their environment, and thus varies with the environment. Animals that do not control their body temperature are **ectotherms**. This group has been called cold-blooded, but the term may not apply to an animal in the desert with a very warm body temperature. In contrast to ectotherms, which rely on external temperatures to set their body temperatures, **poikilotherms** are animals with constantly varying internal temperatures. An animal that maintains a constant body temperature in the face of environmental changes is called an **endotherm**. Endotherms are animals that rely on internal sources for body temperature but which can exhibit extremes in temperature. These animals are able to maintain a level of activity at cooler temperature, whereas an ectotherm cannot.

Heat can be exchanged between an animal and its environment through four mechanisms: radiation, evaporation, convection, and conduction (**Figure 25.50**). Radiation is the emission of electromagnetic “heat” waves. Heat comes from the sun in this manner and radiates from dry skin the same way. Heat can be removed with liquid from a surface during evaporation. This occurs when a mammal sweats. Convection currents of air remove heat from the surface of dry skin as the air passes over it. Heat will be conducted from one surface to another during direct contact with the surfaces, such as an animal resting on a warm rock.



(a) Radiation (b) Evaporation

(c) Convection (d) Conduction

Figure 25.50 Heat can be exchanged by four mechanisms: (a) radiation, (b) evaporation, (c) convection, or (d) conduction. (credit b: modification of work by “Kullez”/Flickr; credit c: modification of work by Chad Rosenthal; credit d: modification of work by “stacey.d”/Flickr)

Neural Control of Thermoregulation

The nervous system is important to **thermoregulation**, as illustrated in **Figure 25.50**. The processes of homeostasis and temperature control are centered in the hypothalamus of the advanced animal brain.

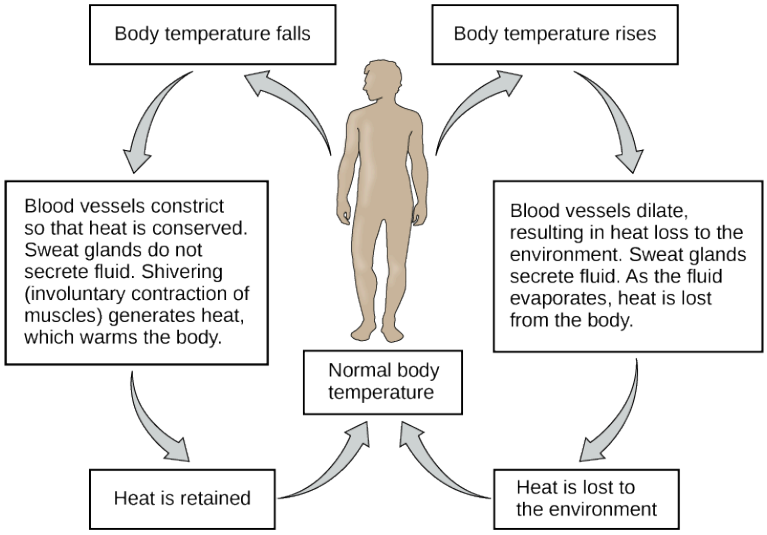


Figure 25.51 The body is able to regulate temperature in response to signals from the nervous system.

The hypothalamus maintains the set point for body temperature through reflexes that cause vasodilation and sweating when the body is too warm, or vasoconstriction and shivering when the body is too cold. It responds to chemicals from the body. When a bacterium is destroyed by phagocytic leukocytes, chemicals called endogenous pyrogens (pyr=fire and genein=to produce) are released into the blood. These chemicals circulate to the hypothalamus and reset the thermostat. This allows the body’s temperature to increase in what is commonly called a fever. An increase in body temperature causes iron to be conserved, inhibiting bacterial division since iron is an essential nutrient for bacteria. An increase in body heat also increases the activity of the animal’s enzymes and protective cells while inhibiting the enzymes and activity of the invading

microorganisms. Finally, heat itself may also kill the pathogen. Thus, a fever that was once thought to be a complication of an infection, is now understood to be a normal defense mechanism.

26 | DIGESTION AND NUTRITION

26.1 | Digestive Systems

> “If you were to open up a baby's head - and I am not for a moment suggesting that you should - you would find nothing but an enormous drool gland.”

- Dave Barry

The salivary glands (aka "drool glands") of vertebrates are just one part of the elaborate and integrated organ system we call the digestive system. The digestive system allows organisms, such as us, to 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 complexity of form and function, their digestive systems (including the drool glands) 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 26.1** 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).

(a) (b)

Figure 26.1 Herbivores, like this (a) American Bison and (b) Pipevine Swallowtail caterpillar, eat primarily plant material. (photo credit: David A. Rintoul)

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 26.2a** and tigers are examples of vertebrate carnivores, as are snakes and sharks, while invertebrate carnivores include sea stars, spiders, and ladybugs, shown in **Figure 26.2b**. 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 26.2 Carnivores like the (a) lion eat primarily meat. The (b) ambush bug is also a carnivore that consumes small insects such as flies. (credit a: modification of work by Kevin Pluck; credit b: David A. Rintoul)

Omnivores are animals that eat both plant- and animal-derived food. In Latin, omnivore means to eat everything. Humans, bears (shown in **Figure 26.3a**), and chickens are example of vertebrate omnivores; invertebrate omnivores include cockroaches and crayfish (shown in **Figure 26.3b**).

(a) (b)

Figure 26.3 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)

Types of Digestive Systems

There re two types of digestive systems: incomplete and complete. An **incomplete digestive system** has only one opening to the digestive tract **Figure 26.4a**. Ingested food and excreted waste products pass through the same opening. Incomplete digestive systems are found in Cnidarians and Flatworms. The other type is, naturally, called a **complete digestive system**, and this type has two openings **Figure 26.4b**. Food is ingested through one opening, and waste products are excreted through a separate opening. This type of digestive system is found in all other phyla, including vertebrates.

The advantage of a complete digestive system is that it allows animals to feed continuously, without waiting for the residues of the previous meal to be released from the digestive tract. It also allows specialization of regions of the digestive tract, which means that different food types can be digested more efficiently. A particular region along the digestive tract, with different physical and chemical conditions, can be optimized for one type of nutrient and other regions can be optimized for efficient metabolism of other types of nutrients.

Invertebrate Digestive Systems

Invertebrates can have incomplete or complete digestive systems, as noted above. An incomplete digestive system has a gastrovascular cavity and only one opening for digestion. Platyhelminthes (flatworms), Ctenophora (comb jellies), and Cnidaria (coral, jelly fish, and sea anemones) have this type of digestive system. Gastrovascular cavities, as shown in **Figure 26.4a**, 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.

A complete digestive system, with an alimentary canal **Figure 26.4b**, 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 a complete digestive system. 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.

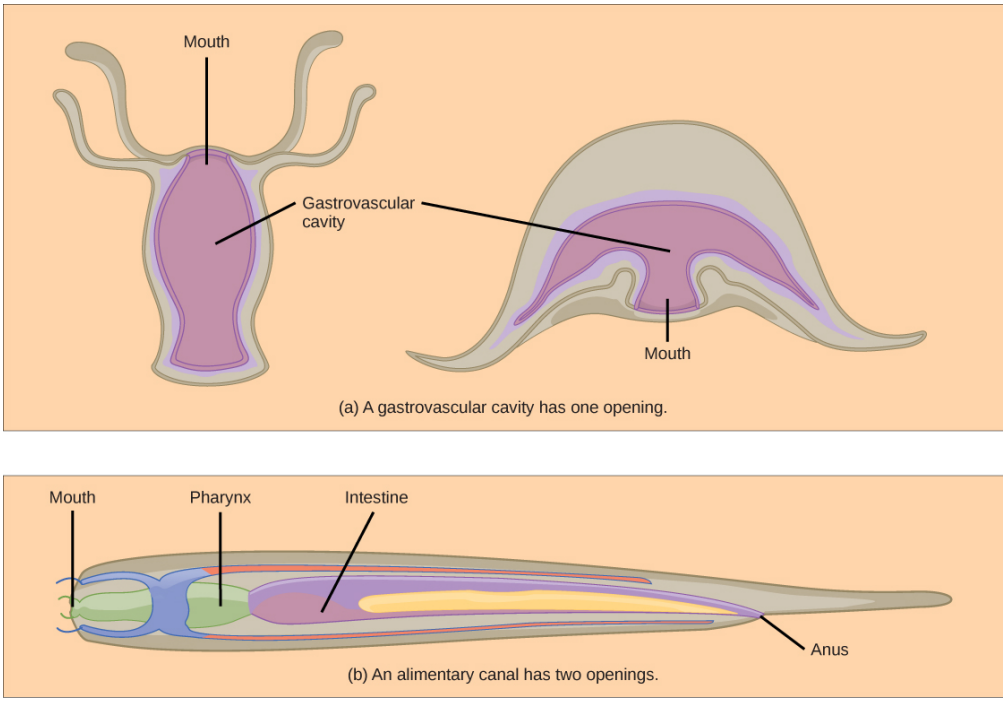


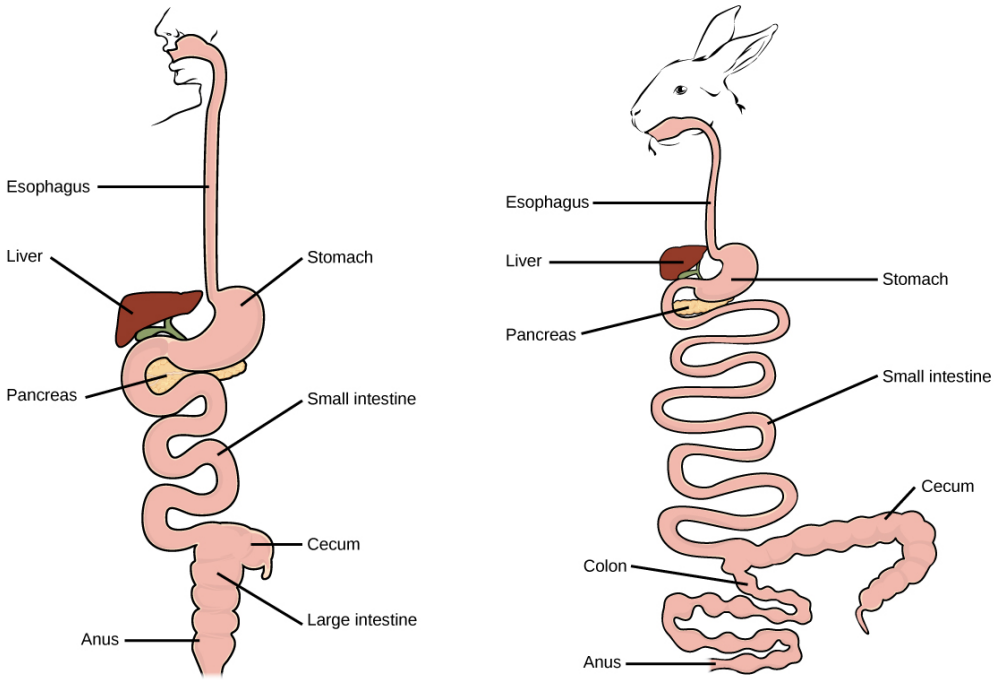
Figure 26.4 (a) A gastrovascular cavity has a single opening through which food is ingested and waste is excreted, as shown in this hydra and in this jellyfish medusa. (b) An alimentary canal has two openings: a mouth for ingesting food, and an anus for eliminating waste, as shown in this nematode.

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 26.5ab**. 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. 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 blood stream 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.



(a) Human digestive system (b) Rabbit digestive system

Figure 26.5 (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

Birds face special challenges when it comes to obtaining nutrition from food. Due to the constraints imposed by having to be lightweight in order to fly, they have lost some adaptations found in their dinosaur ancestors. For example, they do not have teeth, and so their digestive system, shown in **Figure 26.6**, 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.

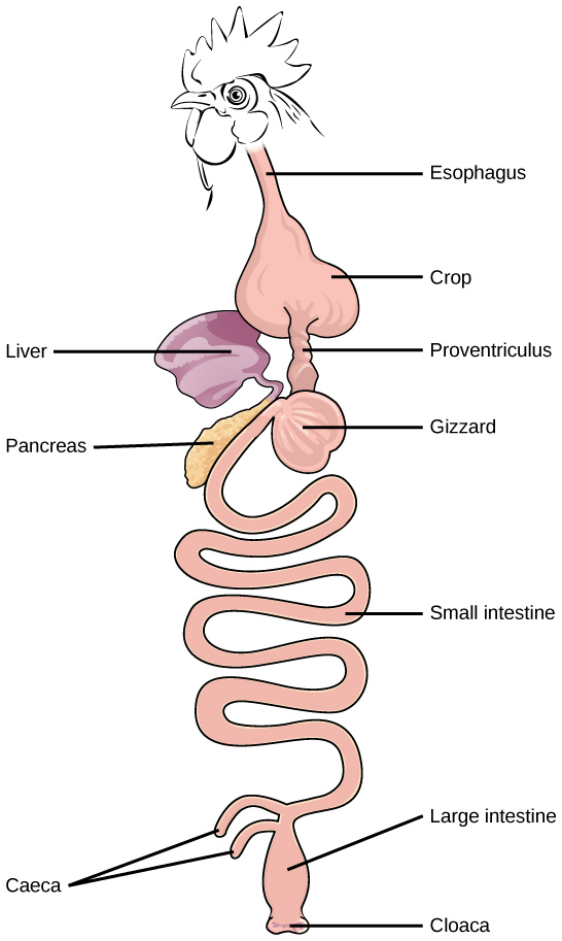


Figure 26.6 The avian esophagus has a pouch, called a crop, which stores food. Food passes from the crop to the first of two stomachs, called the proventriculus, which contains digestive juices that break down food. From the proventriculus, the food enters the second stomach, called the gizzard, which 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.

evolution CONNECTION

Avian Adaptations

Birds have a highly efficient, simplified digestive system. Recent fossil evidence has shown that the evolutionary divergence of birds from other land animals was characterized by streamlining and simplifying the digestive system. Unlike many other animals, birds do not have teeth to chew their food. In place of lips, they have sharp pointy beaks. Instead of teeth, they have a proventricullus, or gizzard, for griding up food. The emergence of these changes seems to coincide with the inclusion of seeds in the bird diet. Seed-eating birds have beaks that are shaped for grabbing seeds and the two-compartment stomach allows for delegation of tasks. Since birds need to remain light in order to fly, passage time in the gut is very short, which means they digest their food very quickly and need to eat often. Contrast this with the ruminants, where the digestion of plant matter takes a very long time and a heavy, water-filled digestive tract. What would you predict would be some characteristics of birds that eat like a cow, ingesting primarily leaves and shoots?

Ruminants

Ruminants are herbivores like cows, sheep, goats, bison, etc. whose entire diet consists of eating large amounts of leaves and shoots. They have evolved digestive systems that help them digest the vast amounts of cellulose in this diet. An interesting

feature of some ruminants mouths is that they do not have upper incisor teeth. They use their lower teeth, tongue and lips to tear and chew their food. From the mouth, the food travels to the esophagus and on to the stomach.

To help digest the large amount of plant material, the stomach of the ruminants is a multi-chambered organ, as illustrated in **Figure 26.7**. The four compartments of the stomach are called the rumen, reticulum, omasum, and abomasum. These chambers contain many microbes that break down cellulose and ferment ingested food. The abomasum is the "true" stomach and is the equivalent of the monogastric stomach chamber where gastric juices are secreted. The four-compartment gastric chamber provides a larger space and the microbial support necessary to digest plant material in ruminants; it is essentially a bacterial fermentation vessel. The fermentation process requires lots of watery fluid, and produces large amounts of gas in the stomach chamber, which must be eliminated. As in other animals, the small intestine plays an important role in nutrient absorption, and the large intestine helps in the elimination of waste.

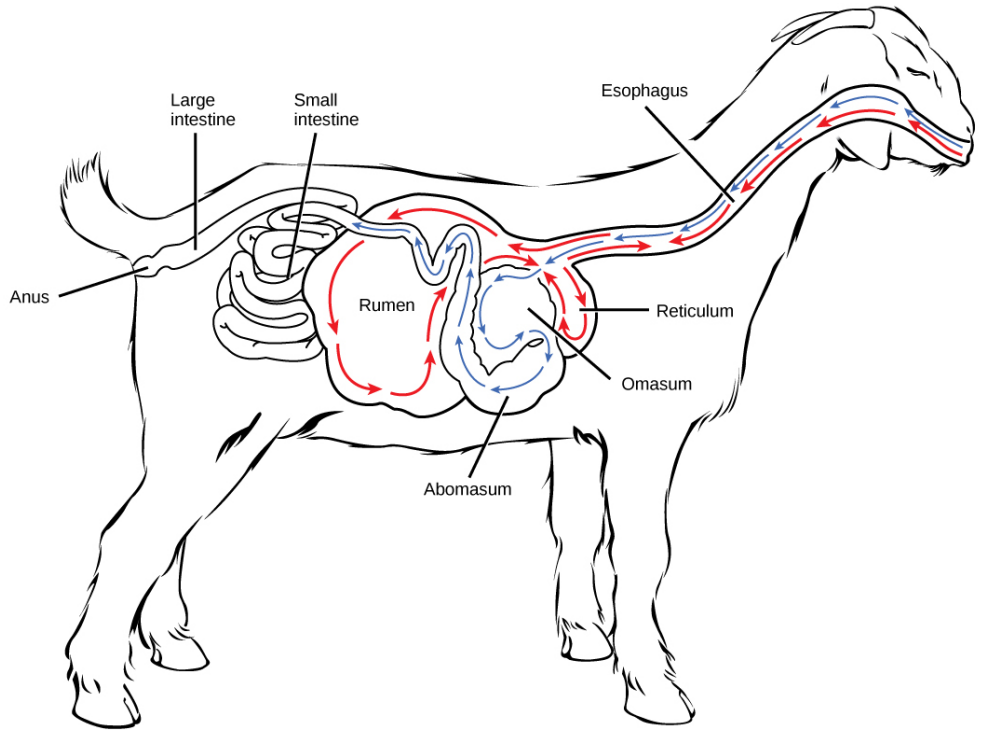


Figure 26.7 Ruminant animals, such as goats and cows, have a four chambered stomach. The first two chambers, the rumen and the reticulum, contain prokaryotes and protists that are able to digest cellulose fiber. The ruminant regurgitates cud from the reticulum, chews it, and swallows it into a third chamber, the omasum, which removes water. The cud then passes onto the fourth chamber, the abomasum, where it is digested by enzymes produced by the ruminant.

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 26.8**. 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 secreted by salivary glands into the mouths of many animals. Saliva contains mucus that moistens food and buffers the pH of the food. Saliva also contains immunoglobulins and lysozymes, which have antibacterial action to reduce tooth decay by inhibiting 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. 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 and lungs.

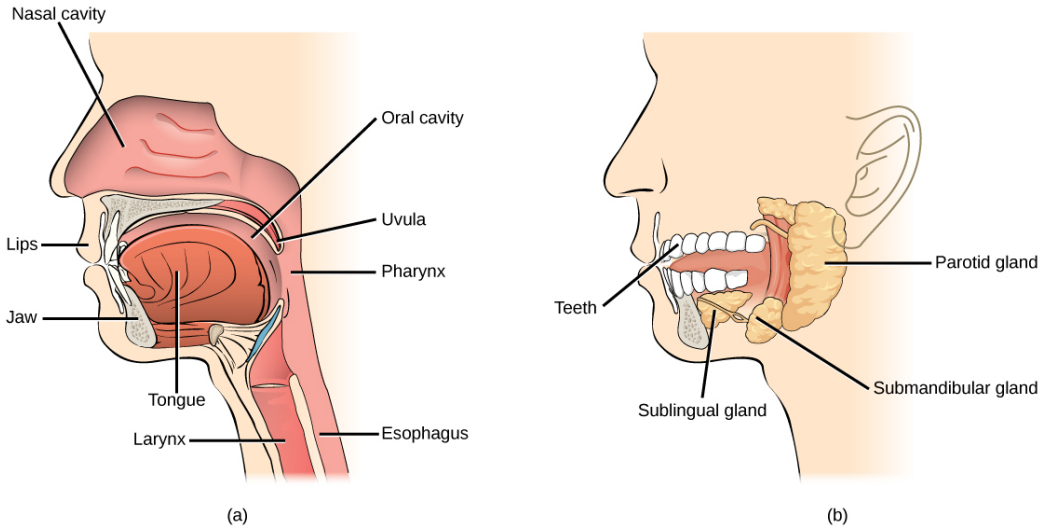


Figure 26.8 Digestion of food begins in the (a) oral cavity. Food is masticated by teeth and moistened by saliva secreted from the (b) salivary glands. Enzymes in the saliva begin to digest starches and fats. With the help of the tongue, the resulting bolus is moved into the esophagus by swallowing. (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 26.9**. The peristalsic 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.

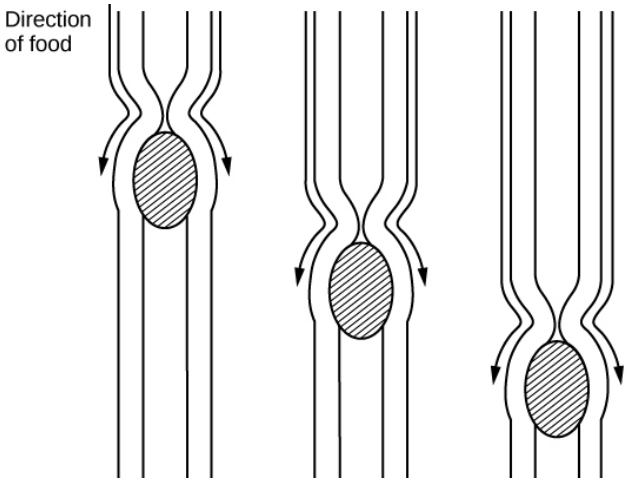


Figure 26.9 The esophagus transfers food from the mouth to the stomach through peristaltic movements.

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 26.10**. The **stomach** is a saclike 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.

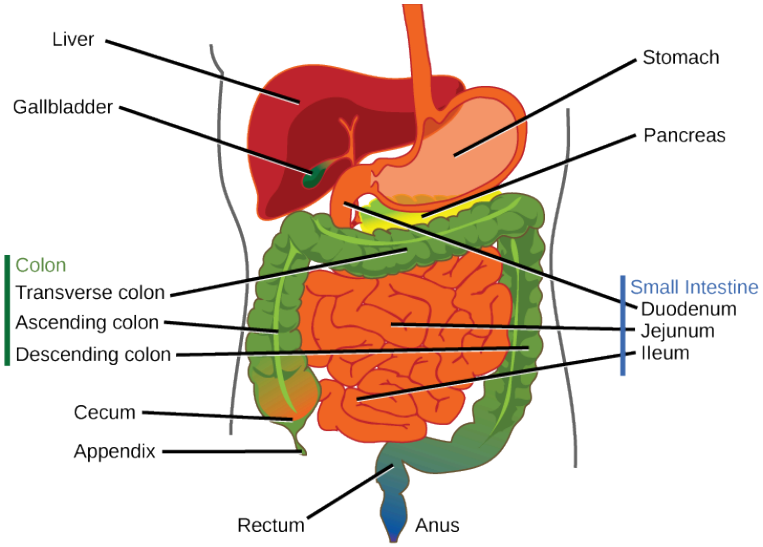


Figure 26.10 The human stomach has an extremely acidic environment where most of the protein gets digested. (credit: modification of work by Mariana Ruiz Villareal)

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** (secreted by cells in the stomach lining) breaks peptide bonds and cleaves proteins into smaller polypeptides. 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. 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 completely mixes the stomach contents about every 20 minutes. The partially digested food and gastric juice mixture is 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. Firstly, the enzyme pepsin is synthesized in an inactive form (pepsinogen), which is activated by the acid and other proteases in the stomach contents. This protects the cells which secrete pepsin, because pepsinogen does not have the full 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 26.11**, 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. Most absorbed nutrients (sugars, amino acids and nucleotides) 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. Fatty acids, resulting from digestion of fat in the small intestine, do not enter the blood stream directly, but are taken up by the lymphatic system in the small intestine. These are eventually delivered to the blood via the thoracic duct, to be metabolized by the liver as is the case for the other nutrients.

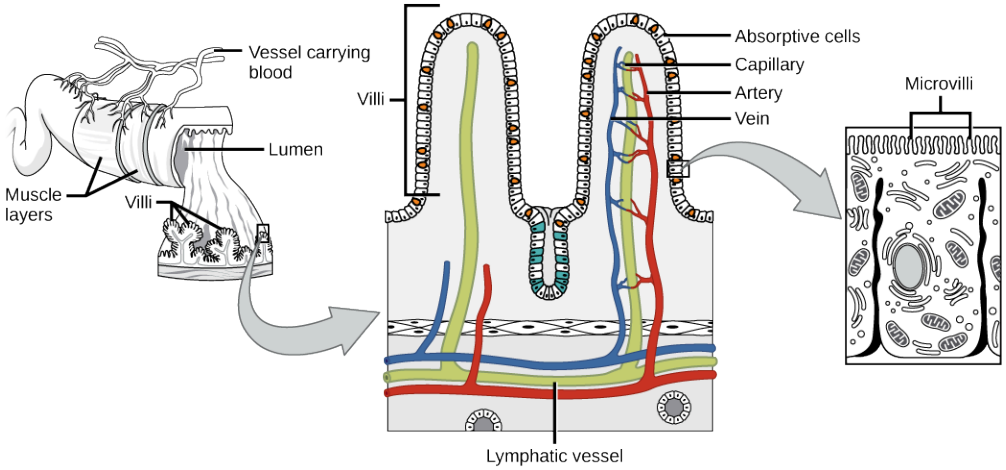


Figure 26.11 Villi are folds on the small intestine lining that increase the surface area to facilitate the absorption of nutrients.

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 26.10**. 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 into the lymphatic system also takes place in the duodenum.

The second part of the small intestine is called the **jejunum**, shown in **Figure 26.10**. 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 26.10** is the last part of the small intestine and here the bile salts and vitamins are absorbed into blood stream. 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 26.12**, 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 cecum and colon are home to many trillions of 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.

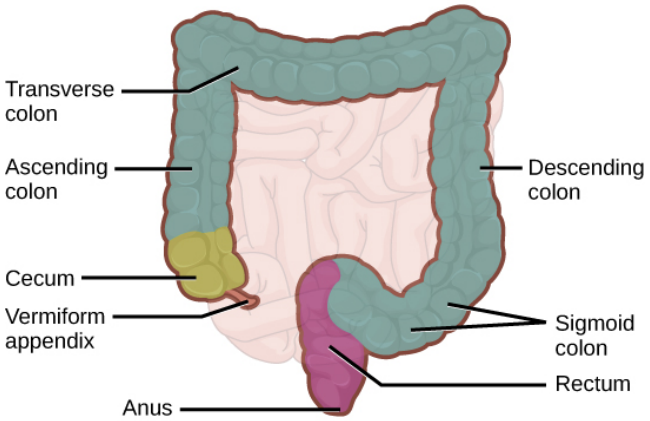


Figure 26.12 The large intestine reabsorbs water from undigested food and stores waste material until it is eliminated.

Rectum and Anus

The **rectum** is the terminal end of the large intestine, as shown in **Figure 26.12**. 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 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.

26.2 | Digestive System Processes

Introduction

> “ Every creature has its own food, and an appropriate alchemist with the task of dividing it ... The alchemist takes the food and changes it into a tincture which he sends through the body to become blood and flesh. This alchemist dwells in the stomach where he cooks and works. The man eats a piece of meat, in which is both bad and good. When the meat reaches the stomach, there is the alchemist who divides it. What does not belong to health he casts away to a special place, and sends the good wherever it is

needed.”

Philippus Aureolus Paracelsus, in *Volumen Medicinae Paramirum*, c. 1520

Obtaining nutrition and energy from food is a multi-step process that, contrary to the thinking of Paracelsus, does not involve alchemy or an alchemist. Many physical and biochemical processes are involved in digestion of food, and it is also a highly regulated 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 polymeric molecules found in intact food cannot pass through plasma membranes. So these polymers need to be broken into smaller monomers so that animal cells can absorb and metabolize them to produce energy. 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 a 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 molecules. It is important to break down macromolecules into smaller monomers that are of suitable size for absorption across the digestive epithelium. Large, complex molecules (e.g., proteins, polysaccharides, nucleic acids, and lipids) must be hydrolyzed into monomers before they can be absorbed by the digestive epithelial cells. If this terminology seems a bit crazy, you may want to review hydrolysis reactions **the section called “Hydrolysis”** before proceeding. Different organs play specific roles in the digestive process. The animal diet needs carbohydrates, protein, nucleic acids, and fat, as well as vitamins and inorganic components for nutritional balance. We will briefly discuss digestion and absorption of some of these in the sections below.

Carbohydrates

The digestion of carbohydrates begins in the mouth. The salivary enzyme amylase begins the breakdown of food starches into maltose, a disaccharide. No significant further digestion of carbohydrates takes place in the stomach. The esophagus produces no digestive enzymes but does produce mucous for lubrication. The acidic environment in the stomach inhibits the action of the salivary amylase enzyme.

The next step of carbohydrate digestion takes place in the duodenum. The chyme from the stomach enters the duodenum and mixes with the digestive secretions from the pancreas, liver, and gallbladder. Pancreatic juices also contain an amylase enzyme, 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 cells lining the small intestine. 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 (e.g., glucose and fructose) thus produced are absorbed by the intestinal cells and transported into the bloodstream. The steps in carbohydrate digestion are summarized in **Figure 26.13** and **Table 26.1**.

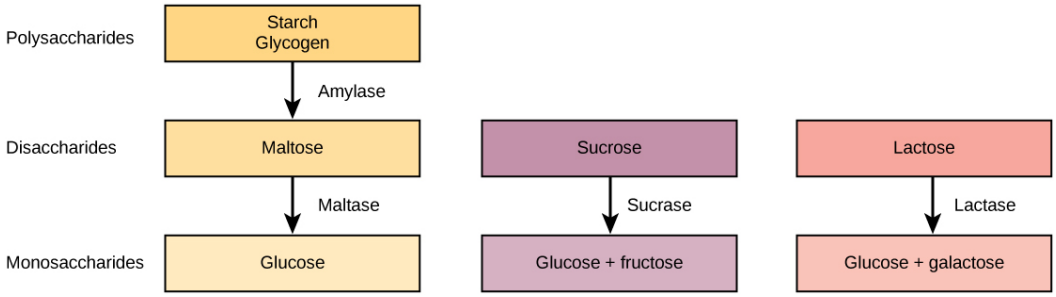


Figure 26.13 Digestion of carbohydrates is performed by several enzymes. Starch and glycogen are broken down into glucose by amylase and maltase. Sucrose (table sugar) and lactose (milk sugar) are broken down by sucrase and lactase, respectively.

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)

Table 26.1

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. 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 26.14** and **Table 26.2**.

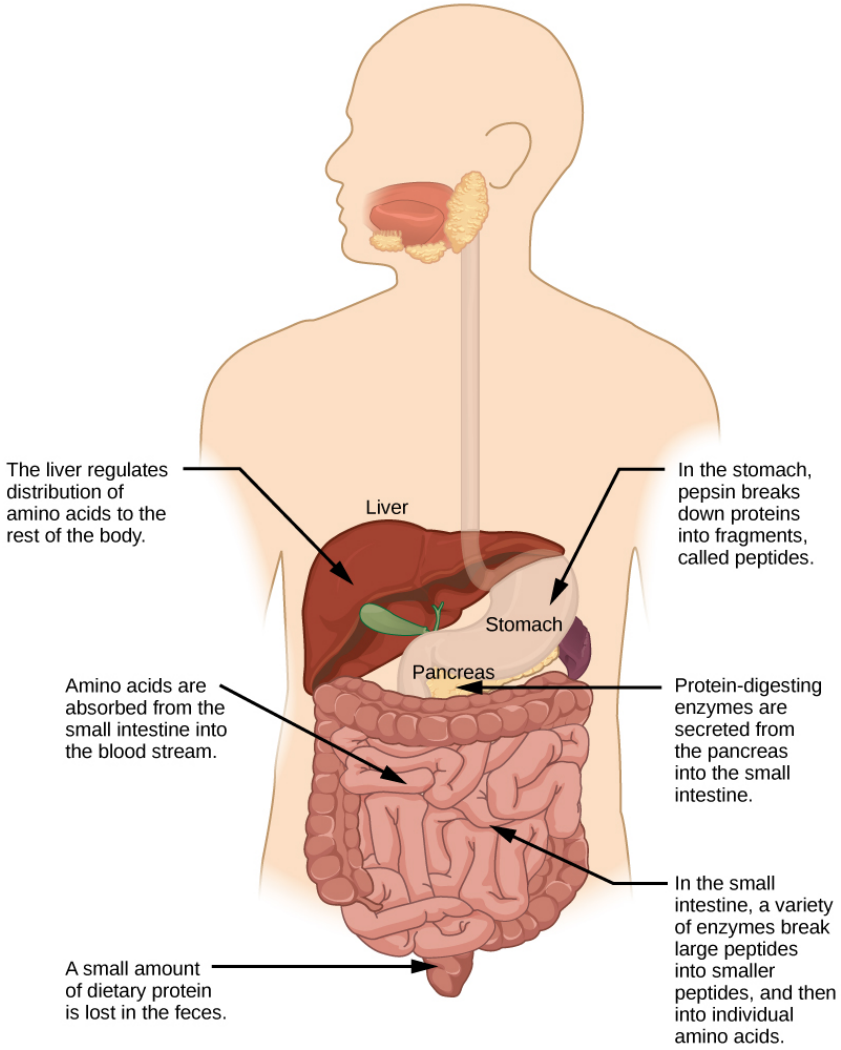


Figure 26.14 Protein digestion is a multistep process that begins in the stomach and continues through the intestines.

Digestion of Protein

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 peptides
Aminopeptidase Dipeptidase	Lining of intestine	Small intestine	Peptides	Amino acids

Table 26.2

Lipids

The bulk of lipid digestion occurs in the small intestine, via the action of pancreatic lipase. When chyme enters the duodenum, it triggers a hormonal response resulting in 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 physical process in which large lipid globules are dispersed into several small lipid globules. Lipids are hydrophobic substances: in the presence of water, they will aggregate to form large globules to minimize exposure to water. These small globules have a larger surface-to-volume ratio and thus an increased surface area for the lipases to interact with. 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.

By forming an emulsion, bile salts increase the available surface area of the lipid particles significantly. The pancreatic lipases can then act on the lipids more efficiently and digest them, as detailed in **Figure 26.15**. Lipases break down the dietary triglycerides into fatty acids and monoglycerides (one fatty acid attached to a glycerol molecule). These molecules can pass through the plasma membrane of the cell and enter the epithelial cells of the intestinal lining. Lipase products (fatty acids and monoglycerides) pass through the intestinal cells where they are reassembled into triglycerides, and then are combined with proteins to form large fatty complexes 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 via the thoracic duct on their way to the liver.

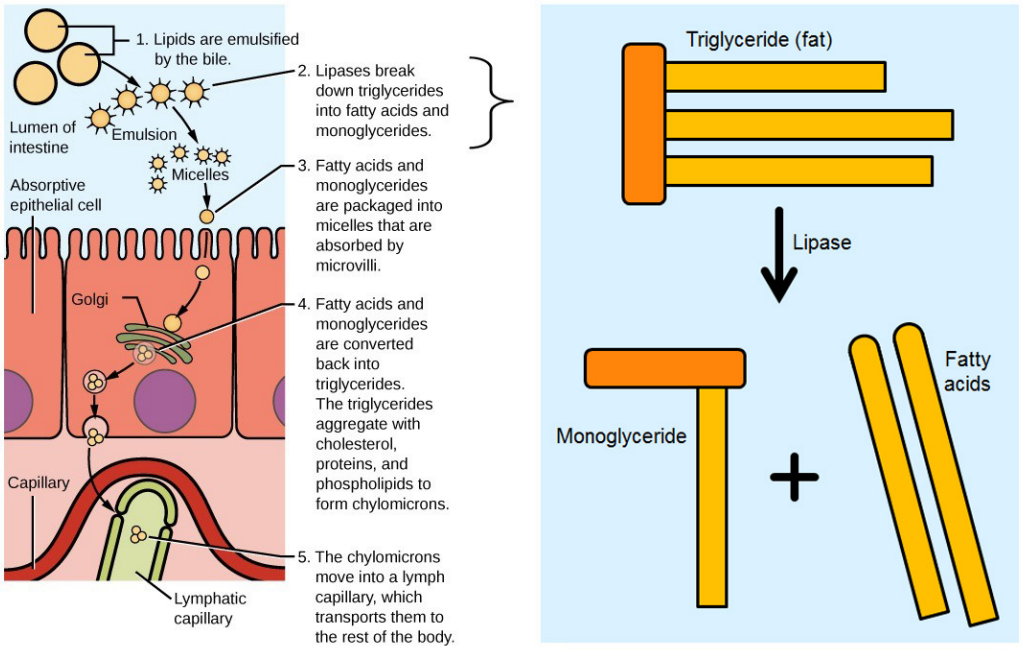


Figure 26.15 Lipids are digested and absorbed in the small intestine.

Summary of Digestion

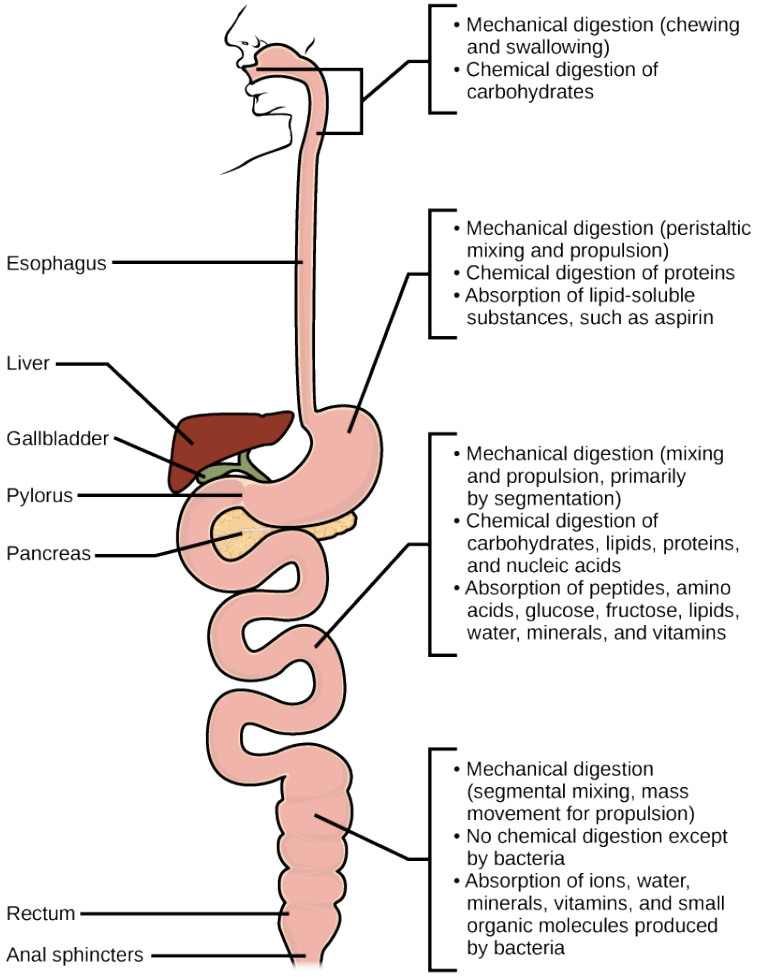


Figure 26.16 Mechanical and chemical digestion of food takes place in many steps, beginning in the mouth and ending in the rectum.

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.

26.3 | Digestive System Regulation

Introduction

"The digestive canal is in its task a complete chemical factory. The raw material passes through a long series of institutions in which it is subjected to certain mechanical and, mainly, chemical processing, and then, through innumerable side-streets, it is brought into the depot of the body. Aside from this basic series of

institutions, along which the raw material moves, there is a series of lateral chemical manufactories, which prepare certain reagents for the appropriate processing of the raw material.”

Ivan Petrovich Pavlov, Speech to the Society of Russian Physicians, Dec. 1874

Pavlov's pioneering work, showing that dogs can associate the ringing of a bell with the imminent delivery of food, pointed the way toward our current understanding of the coordinated responses that regulate the digestive system. The brain is the control center for the sensation of hunger and satiety. The functions of the digestive system are regulated through neural and hormonal responses.

Neural Responses to Food

In reaction to the smell, sight, or thought of food, like that shown in **Figure 26.17**, the first hormonal response is that of salivation. The salivary glands secrete more saliva in response to the stimulus presented by food in preparation for digestion. Simultaneously, the stomach begins to produce hydrochloric acid to digest the food. Recall that the peristaltic movements of the esophagus and other organs of the digestive tract are under the control of the brain. The brain prepares these muscles for movement as well. When the stomach is full, the part of the brain that detects satiety signals fullness. There are three overlapping phases of gastric control—the cephalic phase, the gastric phase, and the intestinal phase—each requires many enzymes and is under neural control as well.

Figure 26.17 Seeing a plate of food triggers the secretion of saliva in the mouth and the production of HCL in the stomach. (credit: Kelly Bailey)

Digestive Phases

The response to food begins even before food enters the mouth. The first phase of ingestion, called the cephalic phase, is controlled by the neural response to the stimulus provided by food. All aspects—such as sight, sense, and smell—trigger the neural responses resulting in salivation and secretion of gastric juices. The gastric and salivary secretion in the cephalic phase can also take place due to the thought of food. Right now, if you think about a piece of chocolate or a crispy potato chip, the increase in salivation is a cephalic phase response to the thought. The central nervous system prepares the stomach to receive food.

The gastric phase begins once the food arrives in the stomach. It builds on the stimulation provided during the cephalic phase. Gastric acids and enzymes process the ingested materials. The gastric phase is stimulated by (1) distension of the stomach, (2) a decrease in the pH of the gastric contents, and (3) the presence of undigested material. This phase consists of local, hormonal, and neural responses. These responses stimulate secretions and powerful contractions.

The intestinal phase begins when chyme enters the small intestine triggering digestive secretions. This phase controls the rate of gastric emptying. In addition to gastric emptying, when chyme enters the small intestine, it triggers other hormonal and neural events that coordinate the activities of the intestinal tract, pancreas, liver, and gallbladder.

Hormonal Responses to Food

The **endocrine system** controls the response of the various glands in the body and the release of hormones at the appropriate times.

One of the important factors under hormonal control is the stomach acid environment. During the gastric phase, the hormone **gastrin** is secreted by G cells in the stomach in response to the presence of proteins in the stomach contents. Gastrin stimulates the release of hydrochloric acid (HCl), which aids in the digestion of the proteins. However, when the stomach is emptied, the acidic environment no longer needs to be maintained, and a hormone called somatostatin stops the release of hydrochloric acid. This is a good example of a negative feedback system: proteins in the stomach cause a response that results in elimination of proteins from the stomach contents, via the actions of two counteracting hormones.

In the duodenum, digestive secretions from the liver, pancreas, and gallbladder play an important role in digesting chyme during the intestinal phase. In order to neutralize the acidic chyme, a hormone called **secretin** stimulates the pancreas to produce alkaline bicarbonate solution and deliver it to the duodenum. Secretin acts in tandem with another hormone called **cholecystokinin** (CCK). Not only does CCK stimulate the pancreas to produce the requisite pancreatic juices, it also stimulates the gallbladder to release bile into the duodenum. This is another negative feedback loop; what parameters are being sensed and regulated?

Another level of hormonal control occurs in response to the composition of food. Foods high in lipids take a long time to digest. A hormone called gastric inhibitory peptide is secreted by the small intestine to slow down the peristaltic movements of the intestine to allow fatty foods more time to be digested and absorbed.

Understanding the hormonal control of the digestive system is an important area of ongoing research. Scientists are exploring the role of each hormone in the digestive process and developing ways to target these hormones. Advances could lead to knowledge that may help to battle the obesity epidemic.

26.4 | Nutrition

> “Tell me what you eat: I will tell you what you are.”

Jean-Anthelme Brillat-Savarin, *The Philosopher in the Kitchen*, Aphorism IV, 1825

Brillat-Savarin's aphorism may not be entirely true, but it holds a grain of truth. 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, and reproduction; collectively these processes are called biosynthesis. The diet is also the source of materials for ATP production in the cells. 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 **Figure 26.18**

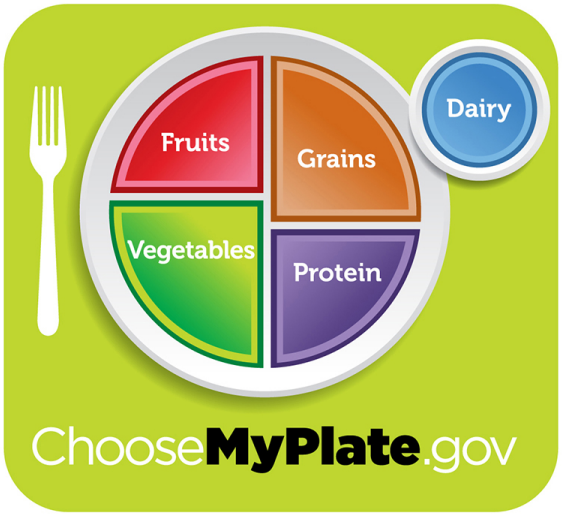


Figure 26.18 For humans, a balanced diet includes fruits, vegetables, grains, and protein. (credit: USDA)

everyday CONNECTION

Let's Move! Campaign

Obesity is a growing epidemic and the rate of obesity among children is rapidly rising in the United States. To combat childhood obesity and ensure that children get a healthy start in life, first lady Michelle Obama has launched the Let's Move! campaign. The goal of this campaign is to educate parents and caregivers on providing healthy nutrition and encouraging active lifestyles to future generations. This program aims to involve the entire community, including parents, teachers, and healthcare providers to ensure that children have access to healthy foods—more fruits, vegetables, and whole grains—and consume fewer calories from processed foods. Another goal is to ensure that children get physical activity. With the increase in television viewing and stationary pursuits such as video games, sedentary lifestyles have become the norm. Learn more at www.letsmove.gov.

Organic Precursors

The organic molecules required for building cellular material and tissues must come from food. Carbohydrates or sugars are the primary source 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 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 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. Absence or low levels of vitamins can have a dramatic effect on health, as outlined in Table 26.3 and Table 26.4. Both fat-soluble and water-soluble vitamins must be obtained from food. **Minerals**, listed in Table 26.5, 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 26.6.

Water-soluble Essential Vitamins

Vitamin	Function	Deficiencies Can Lead To	Sources
Vitamin B_1 (Thiamine)	Needed by the body to process lipids, proteins, and carbohydrates Coenzyme removes CO_2 from organic compounds	Muscle weakness, Beriberi: reduced heart function, CNS problems	Milk, meat, dried beans, whole grains
Vitamin B_2 (Riboflavin)	Takes an active role in metabolism, aiding in the conversion of food to energy (FAD and FMN)	Cracks or sores on the outer surface of the lips (cheliosis); inflammation and redness of the tongue; moist, scaly skin inflammation (seborrheic dermatitis)	Meat, eggs, enriched grains, vegetables
Vitamin B_3 (Niacin)	Used by the body to release energy from carbohydrates and to process alcohol; required for the synthesis of sex hormones; component of coenzyme NAD^+ and $NADP^+$	Pellagra, which can result in dermatitis, diarrhea, dementia, and death	Meat, eggs, grains, nuts, potatoes
Vitamin B_5 (Pantothenic acid)	Assists in producing energy from foods (lipids, in particular); component of coenzyme A	Fatigue, poor coordination, retarded growth, numbness, tingling of hands and feet	Meat, whole grains, milk, fruits, vegetables
Vitamin B_6 (Pyridoxine)	The principal vitamin for processing amino acids and lipids; also helps convert nutrients into energy	Irritability, depression, confusion, mouth sores or ulcers, anemia, muscular twitching	Meat, dairy products, whole grains, orange juice
Vitamin B_7 (Biotin)	Used in energy and amino acid metabolism, fat synthesis, and fat breakdown; helps the body use blood sugar	Hair loss, dermatitis, depression, numbness and tingling in the extremities; neuromuscular disorders	Meat, eggs, legumes and other vegetables
Vitamin B_9 (Folic acid)	Assists the normal development of cells, especially during fetal development; helps metabolize nucleic and amino acids	Deficiency during pregnancy is associated with birth defects, such as neural tube defects and anemia	Leafy green vegetables, whole wheat, fruits, nuts, legumes
Vitamin B_{12} (Cobalamin)	Maintains healthy nervous system and assists with blood cell formation; coenzyme in nucleic acid metabolism	Anemia, neurological disorders, numbness, loss of balance	Meat, eggs, animal products
Vitamin C (Ascorbic acid)	Helps maintain connective tissue: bone, cartilage, and dentin; boosts the immune system	Scurvy, which results in bleeding, hair and tooth loss; joint pain and swelling; delayed wound healing	Citrus fruits, broccoli, tomatoes, red sweet bell peppers

Table 26.3

Fat-soluble Essential Vitamins

Vitamin	Function	Deficiencies Can Lead To	Sources
Vitamin A (Retinol)	Critical to the development of bones, teeth, and skin; helps maintain eyesight, enhances the immune system, fetal development, gene expression	Night-blindness, skin disorders, impaired immunity	Dark green leafy vegetables, yellow-orange vegetables fruits, milk, butter
Vitamin D	Critical for calcium absorption for bone development and strength; maintains a stable nervous system; maintains a normal and strong heartbeat; helps in blood clotting	Rickets, osteomalacia, immunity	Cod liver oil, milk, egg yolk
Vitamin E (Tocopherol)	Lessens oxidative damage of cells,and prevents lung damage from pollutants; vital to the immune system	Deficiency is rare; anemia, nervous system degeneration	Wheat germ oil, unrefined vegetable oils, nuts, seeds, grains
Vitamin K (Phylloquinone)	Essential to blood clotting	Bleeding and easy bruising	Leafy green vegetables, tea

Table 26.4

Figure 26.19 A healthy diet should include a variety of foods to ensure that needs for essential nutrients are met. (credit: Keith Weller, USDA ARS)

Minerals and Their Function in the Human Body

Mineral	Function	Deficiencies Can Lead To	Sources
*Calcium	Needed for muscle and neuron function; heart health; builds bone and supports synthesis and function of blood cells; nerve function	Osteoporosis, rickets, muscle spasms, impaired growth	Milk, yogurt, fish, green leafy vegetables, legumes

Table 26.5

Minerals and Their Function in the Human Body

Mineral	Function	Deficiencies Can Lead To	Sources
*Chlorine	Needed for production of hydrochloric acid (HCl) in the stomach and nerve function; osmotic balance	Muscle cramps, mood disturbances, reduced appetite	Table salt
Copper (trace amounts)	Required component of many redox enzymes, including cytochrome c oxidase; cofactor for hemoglobin synthesis	Copper deficiency is rare	Liver, oysters, cocoa, chocolate, sesame, nuts
Iodine	Required for the synthesis of thyroid hormones	Goiter	Seafood, iodized salt, dairy products
Iron	Required for many proteins and enzymes, notably hemoglobin, to prevent anemia	Anemia, which causes poor concentration, fatigue, and poor immune function	Red meat, leafy green vegetables, fish (tuna, salmon), eggs, dried fruits, beans, whole grains
*Magnesium	Required co-factor for ATP formation; bone formation; normal membrane functions; muscle function	Mood disturbances, muscle spasms	Whole grains, leafy green vegetables
Manganese (trace amounts)	A cofactor in enzyme functions; trace amounts are required	Manganese deficiency is rare	Common in most foods
Molybdenum (trace amounts)	Acts as a cofactor for three essential enzymes in humans: sulfite oxidase, xanthine oxidase, and aldehyde oxidase	Molybdenum deficiency is rare	
*Phosphorus	A component of bones and teeth; helps regulate acid-base balance; nucleotide synthesis	Weakness, bone abnormalities, calcium loss	Milk, hard cheese, whole grains, meats
*Potassium	Vital for muscles, heart, and nerve function	Cardiac rhythm disturbance, muscle weakness	Legumes, potato skin, tomatoes, bananas
Selenium (trace amounts)	A cofactor essential to activity of antioxidant enzymes like glutathione peroxidase; trace amounts are required	Selenium deficiency is rare	Common in most foods
*Sodium	Systemic electrolyte required for many functions; acid-base balance; water balance; nerve function	Muscle cramps, fatigue, reduced appetite	Table salt
Zinc (trace amounts)	Required for several enzymes such as carboxypeptidase, liver alcohol dehydrogenase, and carbonic anhydrase	Anemia, poor wound healing, can lead to short stature	Common in most foods
*Greater than 200mg/day required			

Table 26.5

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.

Table 26.6

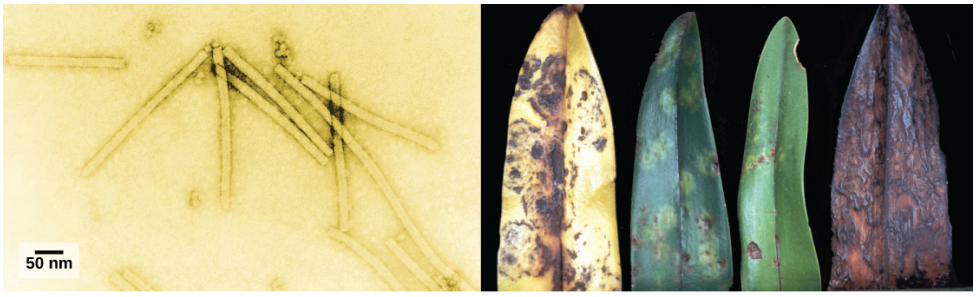
27 | CIRCULATORY, RESPIRATORY AND IMMUNE SYSTEMS

27.1 | Viruses

Introduction

> "That's the salubrious thing about zoonotic diseases: they remind us, as Saint Francis did, that we humans are inseparable from the natural world. In fact, there **is** no "natural world", its a bad and artificial phrase. There is only the world. Humankind is part of that world, as are the ebola viruses, as are the influenzas and the HIVs, as are Marburg and Nipah and SARS, as are chimpanzees and palm civets and Egyptian fruit bats, as is the next murderous virus - the one we haven't yet discovered."

David Quammen, *Spillover: Animal Infections and the Next Human Pandemic*, 2012



(a) (b)

Figure 27.1 (a) The tobacco mosaic virus, seen by transmission electron microscopy, was the first virus to be discovered. (b) The leaves of an infected plant are shown. (credit a: scale-bar data from Matt Russell; credit b: modification of work by USDA, Department of Plant Pathology Archive, North Carolina State University)

No one knows exactly when viruses emerged or from where they came, since viruses do not leave physical evidence in the form of fossils. Modern viruses are thought to be a mosaic of bits and pieces of nucleic acids picked up from various sources along their respective evolutionary paths. Viruses are acellular, parasitic entities that are not classified within any of the three domains because they are not exactly alive. But they do parasitize, evolve, reproduce and co-evolve with other organisms; they inhabit a shadowy world that may not be alive, but is very close to it. They have no plasma membrane, internal organelles, or metabolic processes, and they do not divide. Instead, they infect a host cell and use the host's replication processes to produce progeny virus particles. Viruses infect all forms of organisms including bacteria, archaea, fungi, plants, and animals.

Viruses are diverse. They vary in their structure, their replication methods, and in their target hosts or even host cells.

They infect every type of organism known, from Archaea to Bacteria to Eukaryotes, and are found in every environment. They are also remarkably abundant; it is estimated that each milliliter of sea water contains 10^7 viruses, both DNA and RNA varieties. They are major players in the evolution of the life forms on this planet; genes derived from viruses allowed mammals to develop a placenta, for example.

How Viruses Replicate

Viruses were first discovered after the development of a porcelain filter, called the Chamberland-Pasteur filter, which could remove all bacteria visible under the microscope from any liquid sample. In 1886, Adolph Meyer demonstrated that a disease of tobacco plants, tobacco mosaic disease, could be transferred from a diseased plant to a healthy one through liquid plant extracts. In 1892, Dmitri Ivanowski showed that this disease could be transmitted in this way even after the Chamberland-Pasteur filter had removed all viable bacteria from the extract. Still, it was many years before it was proven that these "filterable" infectious agents were not simply very small bacteria but were a new type of tiny, disease-causing particle.

Virions, single virus particles, are very small, about 20–250 nanometers (1 nanometer = 1/1,000,000 mm); although the recent discovery of entities called Pandoraviruses (approx 1 micrometer, or 1/1,000 mm in diameter) has shaken that paradigm somewhat. Individual virus particles are the infectious form of a virus outside the host cell. Unlike bacteria (which are about 100 times larger), we cannot see most viruses with a light microscope, with the exception of the Pandoraviruses and some large virions of the poxvirus family (**Figure 27.2**).

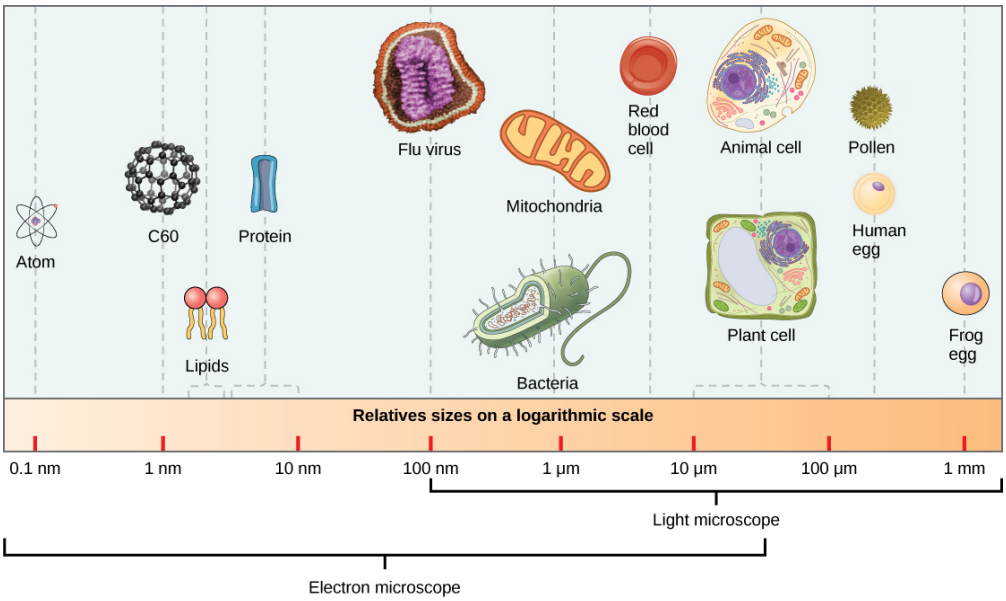


Figure 27.2 The size of a virus is very small relative to the size of cells and organelles.

It was not until the development of the electron microscope in the 1940s that scientists got their first good view of the structure of the tobacco mosaic virus (**Figure 27.1**) and others. The surface structure of virions can be observed by both scanning and transmission electron microscopy, whereas the internal structures of the virus can only be observed in images from a transmission electron microscope (**Figure 27.3**).

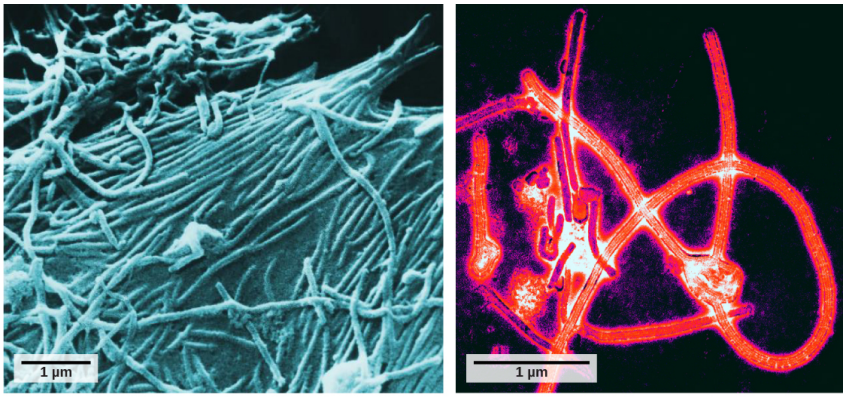


Figure 27.3 The ebola virus is shown here as visualized through (a) a scanning electron micrograph and (b) a transmission electron micrograph. (credit a: modification of work by Cynthia Goldsmith, CDC; credit b: modification of work by Thomas W. Geisbert, Boston University School of Medicine; scale-bar data from Matt Russell)

The use of this technology has allowed for the discovery of many viruses of all types of living organisms. They were initially grouped by shared morphology, meaning their size, shape, and distinguishing structures. Later, groups of viruses were classified by the type of nucleic acid they contained, DNA or RNA, and whether their nucleic acid was single- or double-stranded. More recently, molecular analysis of viral replication cycles has further refined their classification. Currently virus classification begins at the level of Order, and proceeds to species level taxonomy using this scheme. The terms in parentheses are the taxon suffixes for that taxonomic level.

Virus classification

- Order (-virales)
- Family (-viridae)
- Subfamily (-virinae)
- Genus (-virus)
- Species (usually XXXX (disease) virus, e.g., Tobacco Mosaic Virus)

A virion consists of a nucleic-acid core, an outer protein coating, and sometimes an outer envelope made of protein and phospholipids derived from the host cell. The most visible difference between members of viral families is their morphology, which is quite diverse. An interesting feature of viral complexity is that the complexity of the host does not correlate to the complexity of the virion. Some of the most complex virion structures are observed in bacteriophages, viruses that infect the simplest living organisms, bacteria.

Viruses come in many shapes and sizes, but these are consistent and distinct for each viral family (**Figure 27.4**). All virions have a nucleic-acid genome covered by a protective layer of protein, called a capsid. The capsid is made of protein subunits called capsomeres. Some viral capsids are simple polyhedral “spheres,” whereas others are quite complex in structure. The outer structure surrounding the capsid of some viruses is called the viral envelope. All viruses use some sort of glycoprotein to attach to their host cells at molecules on the cell called viral receptors. The virus exploits these cell-surface molecules, which the cell uses for some other purpose, as a way to recognize and infect specific cell types.

The T4 bacteriophage, which infects the *E. coli* bacterium, is among the most complex virions known; T4 has a protein tail structure that the virus uses to attach to the host cell and a head structure that houses its DNA.

Adenovirus, a nonenveloped animal virus that causes respiratory illnesses in humans, uses protein spikes protruding from its capsomeres to attach to the host cell. Nonenveloped viruses also include those that cause polio (poliovirus), plantar warts (papillomavirus), and hepatitis A (hepatitis A virus). Nonenveloped viruses tend to be more robust and more likely to survive under harsh conditions, such as the gut.

Enveloped virions like HIV (human immunodeficiency virus), the causative agent in AIDS (acquired immune deficiency syndrome), consist of nucleic acid (RNA in the case of HIV) and capsid proteins surrounded by a phospholipid bilayer envelope and its associated proteins (**Figure 27.4**). Chicken pox, influenza, and mumps are examples of diseases caused by viruses with envelopes. Because of the fragility of the envelope, nonenveloped viruses are more resistant to changes in temperature, pH, and some disinfectants than enveloped viruses.

Overall, the shape of the virion and the presence or absence of an envelope tells us little about what diseases the viruses may cause or what species they might infect, but is still a useful means to begin viral classification.

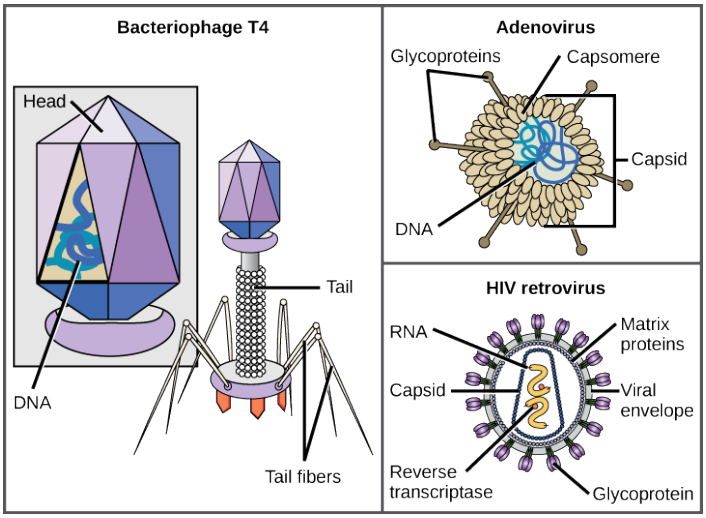


Figure 27.4 Viruses can be complex in shape or relatively simple. This figure shows three relatively complex virions: the bacteriophage T4, with its DNA-containing head group and tail fibers that attach to host cells; adenovirus, which uses spikes from its capsid to bind to the host cells; and HIV, which uses glycoproteins embedded in its envelope to do so. Notice that HIV has proteins called matrix proteins, internal to the envelope, which help stabilize virion shape. HIV is a retrovirus, which means it reverse transcribes its RNA genome into DNA, which is then spliced into the host's DNA. (credit "bacteriophage, adenovirus": modification of work by NCBI, NIH; credit "HIV retrovirus": modification of work by NIAID, NIH)

Unlike all living organisms that use DNA as their genetic material, viruses may use either DNA or RNA as theirs. The virus core contains the genome or total genetic content of the virus. Viral genomes tend to be small compared to bacteria or eukaryotes, containing only those genes that code for proteins the virus cannot get from the host cell. This genetic material may be single-stranded or double-stranded. It may also be linear or circular. While most viruses contain a single segment of nucleic acid, others have genomes that consist of several segments. All of these features are used to help classify viruses into orders, families, etc.

DNA viruses have a DNA core. The viral DNA directs the host cell's replication proteins to synthesize new copies of the viral genome and to transcribe and translate that genome into viral proteins. DNA viruses cause human diseases such as chickenpox, hepatitis B, and some venereal diseases like herpes and genital warts.

RNA viruses contain only RNA in their cores. To replicate their genomes in the host cell, the genomes of RNA viruses encode enzymes not found in host cells. RNA polymerase enzymes are not as stable as DNA polymerases and often make mistakes during transcription. For this reason, mutations, changes in the nucleotide sequence, in RNA viruses occur more frequently than in DNA viruses. This leads to more rapid evolution and change in RNA viruses. For example, the fact that influenza is an RNA virus is one reason a new flu vaccine is needed every year; rapid evolution results in new flu strains being produced constantly in various parts of the world. Human diseases caused by RNA viruses include hepatitis, measles, HIV, common cold virus, Ebola and rabies.

Steps of Virus Infections

Viruses are specialized parasites, usually only infecting one type of cell or one type of organism. A virus must "take over" a cell to replicate. The viral replication cycle can produce dramatic biochemical and structural changes in the host cell, which may cause cell damage. These changes, called cytopathic effects, can change cell functions or even destroy the cell. Some infected cells, such as those infected by the common cold virus (rhinovirus), die through lysis (bursting) or apoptosis (programmed cell death or "cell suicide"), releasing all the progeny virions at once. The symptoms of these viral diseases result from the immune response to the virus, which attempts to control and eliminate the virus from the body, and from cell damage caused by the virus. Many animal viruses, such as HIV (human immunodeficiency virus), leave the infected cells of the immune system by a process known as budding, where virions leave the cell individually. During the budding process, the cell does not undergo lysis and is not immediately killed. However, the damage to the cells that HIV infects may make it impossible for the cells to function as mediators of immunity, even though the cells remain alive for a period of time. Most productive viral infections follow similar steps in the , **virus replication cycle** : attachment, penetration, uncoating, replication, assembly, and release.

A virus attaches to a specific receptor site on the host-plasma membrane through attachment proteins in the capsid or proteins embedded in its envelope. The attachment is specific, and typically a virus will only attach to cells of one or a few species and only certain cell types within those species with the appropriate receptors.

The nucleic acid of bacteriophages is injected directly into the host cell, leaving the capsid outside the cell. Plant and animal viruses can enter their cells through endocytosis, in which the plasma membrane surrounds and engulfs the entire virus. Some enveloped viruses enter the cell when the viral envelope fuses directly with the plasma membrane. Once inside the cell, the viral capsid is degraded and the viral nucleic acid is released, which then becomes available for replication and transcription. Obviously, the naked DNA of a bacteriophage is already available for transcription and replication immediately after being injected into the bacterial cell.

The replication mechanism depends on the viral genome (DNA or RNA). DNA viruses usually use host cell proteins and enzymes to make additional DNA that is then used to copy the genome or be transcribed to messenger RNA (mRNA). The mRNA is then used in protein synthesis. RNA viruses, such as the influenza virus, usually use the RNA as a template for synthesis of viral genomic RNA and mRNA. The viral mRNA is translated into viral enzymes and capsid proteins to assemble new virions (**Figure 27.5**).

The last stage of viral replication is the release of the new virions into the host organism, where they are able to infect adjacent cells and repeat the replication cycle. Some viruses are released when the host cell dies and other viruses can leave infected cells by budding through the membrane without directly killing the cell.

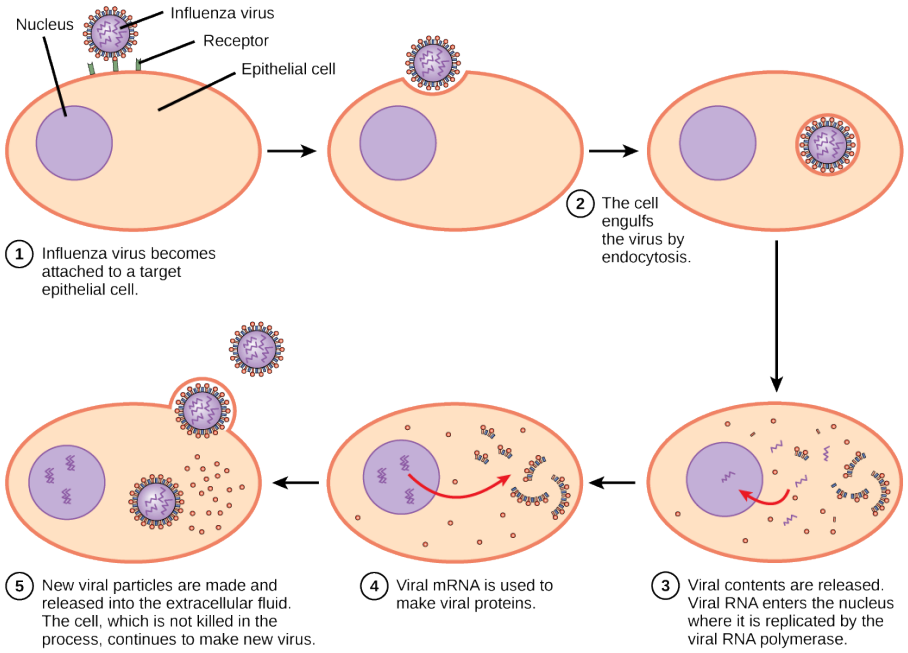


Figure 27.5 In influenza virus infection, glycoproteins attach to a host epithelial cell. As a result, the virus is engulfed. RNA and proteins are made and assembled into new virions.

Lytic and Lysogenic Pathways

Cell death may be immediate or delayed after attachment and penetration by the virus. For example, bacteriophages, viruses that infect bacteria, may or may not kill their host immediately. There are two viral replication strategies; when the virus kills the host cell it is called the **lytic cycle**, and when the virus does not kill the host but replicates when the host replicates it is called the **lysogenic cycle** (**Figure 27.6**).

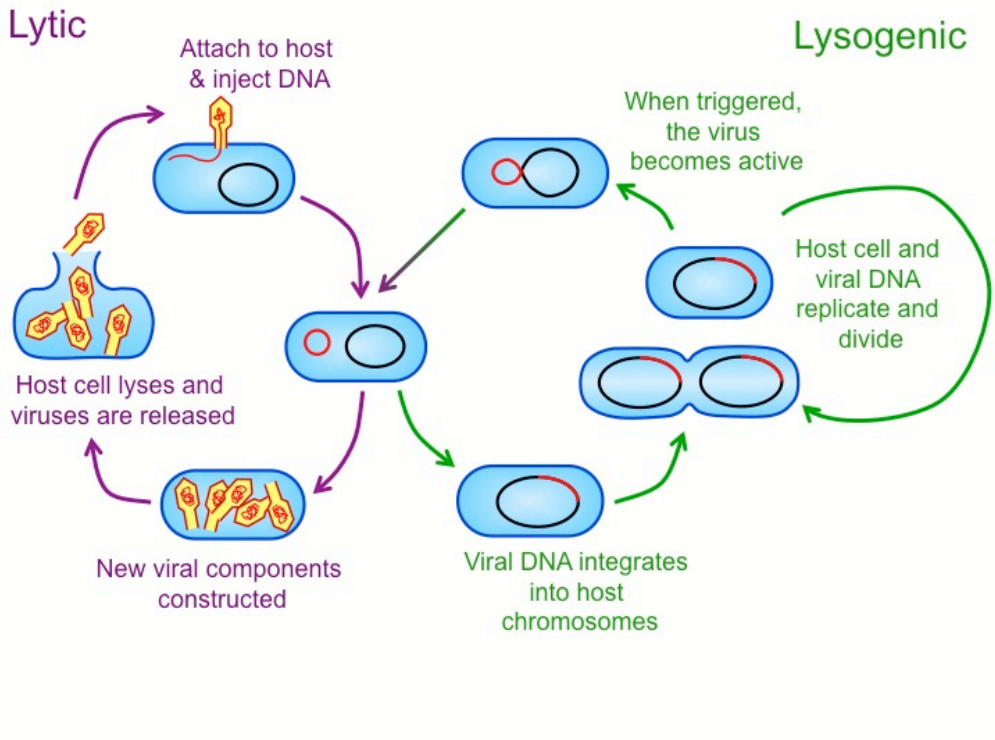


Figure 27.6 The two viral reproductive strategies, the lytic cycle and the lysogenic cycle

Lytic cycle

The lytic cycle causes death of the host cell and the term refers to the last stage of the infection when the cell lyses (breaks open) and releases new virions that were produced within the cell. These new virions can infect healthy cells and the cycle is repeated (**Figure 27.6**).

So why haven't all the bacteria in the world been destroy by bacteriophages? The answer is natural selection of defense mechanisms by bacteria. Mutations of bacterial surface proteins that are not recognized by a particular phage allow the bacteria to survive by preventing attachment. Without going into detail, bacteria have internal defenses that allow them to cut up viral DNA before it can infect the cell. Then one might ask, why hasn't all the bacteriophages in the world gone extinct by not being able to reproduce. Once again, the answer is natural selection. Viruses mutate to bypass the defense mechanisms of the bacteria. This illustrates that the parasite-host relationship is in a constant evolutionary duel.Similar co-evolutionary strategies characterize the interactions of viruses and animals, or viruses and plants.

Lysogenic cycle

There is another reason why bacteria are not extinct because of bacteriophages. Many bacteriophages do not kill their host but coexist within their host, and when this occurs it is called the lysogenic cycle. After penetration, the viral DNA or RNA can either be incorporated into the host DNA, or the viral genome can be a self-replicating entity. Once this occurs, the viral genome is replicated along with the host cell's DNA, but the virus does not destroy the cell as it does in the lytic cycle (**Figure 27.6**). However, at some point the viral genes are turned on and can trigger the virus to enter the lytic cycle and kill the host cell (**Figure 27.6**). Cell starvation or cell damage (e.g. from radiation) may trigger a lysogenic infection to turn into a lytic infection thereby killing the host cell. The next generation of viruses, depending on the host cell condition, can use either of the viral replication strategies, lysogenic or lytic, on the next host.

Viruses and Disease

Viruses cause a variety of diseases in animals, including humans, ranging from the common cold to potentially fatal illnesses like meningitis (**Figure 27.7**). These diseases can be treated by antiviral drugs or by vaccines, but some viruses, such as HIV, are capable of avoiding the immune response and mutating so as to become resistant to antiviral drugs.

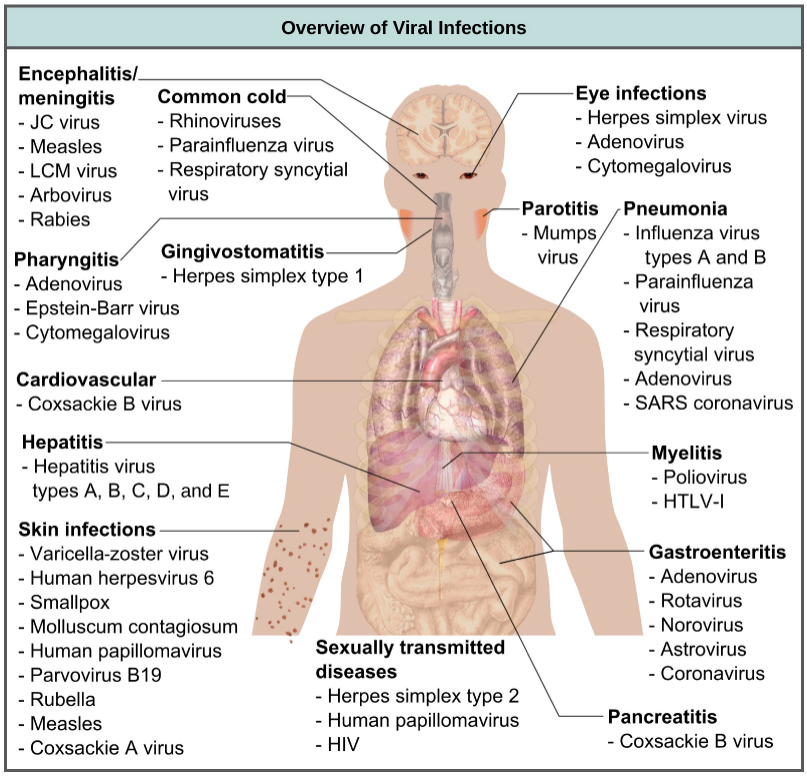


Figure 27.7 Viruses are the cause of dozens of ailments in humans, ranging from mild illnesses to serious diseases. (credit: modification of work by Mikael Häggström)

27.2 | The Circulatory System

Introduction

> "Observation by means of the microscope will reveal more wonderful things than those viewed in regard to mere structure and connection: for while the heart is still beating the contrary (i.e., in opposite directions in the different vessels) movement of the blood is observed in the vessels—though with difficulty—so that the circulation of the blood is clearly exposed."

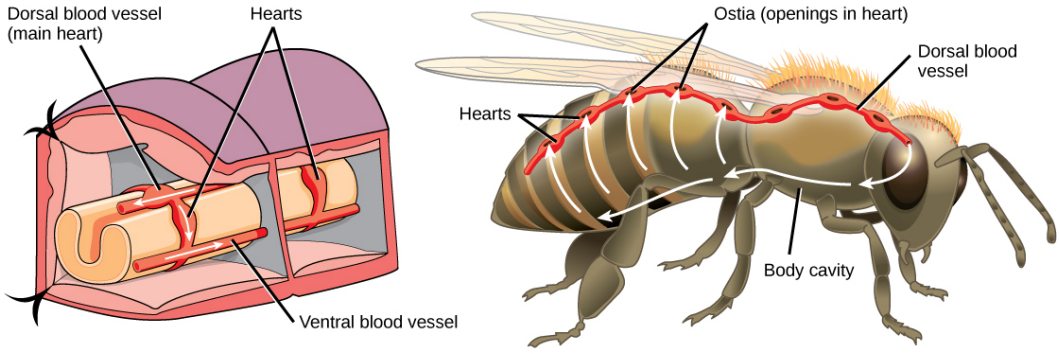
Marcello Malpighi, *De Pulmonibus*, 1661

Malpighi's work (mostly on frogs) outlined the finer microscopic details of circulation, following the work of Harvey, who described the circulatory system at a macroscopic level. In all animals, except a few simple types, the circulatory system is used to transport nutrients and gases through the body. Simple diffusion allows some water, nutrient, waste, and gas exchange into primitive animals that are only a few cell layers thick; however, bulk flow is the only method by which the entire body of larger more complex organisms is accessed.

Circulatory System Architecture

The circulatory system is effectively a network of cylindrical vessels: the arteries, veins, and capillaries that emanate from a pump, the heart. In all vertebrate organisms, as well as some invertebrates, this is a closed system, in which the blood is not free in a cavity. In a **closed circulatory system**, blood is contained inside blood vessels and circulates unidirectionally from

the heart around the systemic circulatory route, then returns to the heart again, as illustrated in **Figure 27.8a**. As opposed to a closed system, arthropods—including insects, crustaceans, and most mollusks—have an open circulatory system, as illustrated in **Figure 27.8b**. In an **open circulatory system**, the fluid is not enclosed in the blood vessels but is pumped into a cavity called a hemocoel; rather than blood, this fluid is called hemolymph because the blood mixes with the interstitial fluid. As the heart beats and the animal moves, the hemolymph circulates around the organs within the body cavity and then reenters the hearts through openings called ostia. This movement allows for gas and nutrient exchange. An open circulatory system does not use as much energy as a closed system to operate or to maintain; however, there is a trade-off with the amount of blood that can be moved to highly metabolically active organs and tissues.

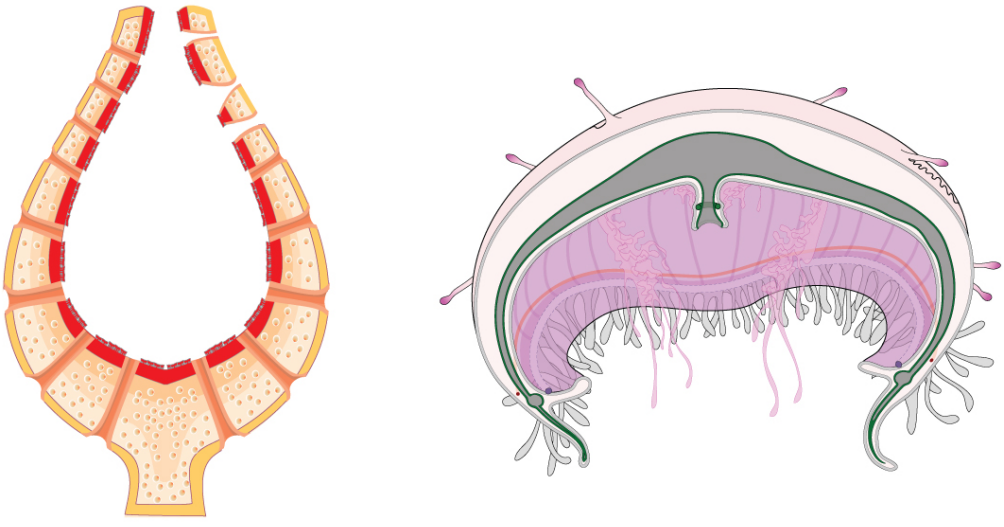


(a) Closed circulatory system (b) Open circulatory system

Figure 27.8 In (a) closed circulatory systems, the heart pumps blood through vessels that are separate from the interstitial fluid of the body. Most vertebrates and some invertebrates, like this annelid earthworm, have a closed circulatory system. In (b) open circulatory systems, a fluid called hemolymph is pumped through a blood vessel that empties into the body cavity. Hemolymph returns to the blood vessel through openings called ostia. Arthropods like this bee and most mollusks have open circulatory systems.

Circulatory System Variation in Animals

The circulatory system varies from simple systems in invertebrates to more complex systems in vertebrates. The simplest animals, such as the sponges (Porifera) and rotifers (Rotifera), do not need a circulatory system because diffusion allows adequate exchange of water, nutrients, and waste, as well as dissolved gases, as shown in **Figure 27.9a**. Organisms that are more complex but still only have two layers of cells in their body plan, such as jellies (Cnidaria) and comb jellies (Ctenophora) also use diffusion through their epidermis and internally through the gastrovascular compartment. Both their internal and external tissues are bathed in an aqueous environment and exchange fluids by diffusion on both sides, as illustrated in **Figure 27.9b**. Exchange of fluids is assisted by the pulsing of the jellyfish body.



(a) Sponge (b) Jellyfish

Figure 27.9 Simple animals consisting of a single cell layer such as the (a) sponge or only a few cell layers such as the (b) jellyfish do not have a circulatory system. Instead, gases, nutrients, and wastes are exchanged by diffusion.

For more complex organisms, diffusion is not efficient for moving gases, nutrients, and waste effectively through the

body; natural selection led to the development of more efficient systems. Most arthropods and many mollusks have open circulatory systems. In an open system, an elongated beating heart pushes the hemolymph through the body, and muscle contractions help to move fluids. The larger more complex crustaceans, including lobsters, have developed arterial-like vessels to push blood through their bodies, and the most active mollusks, such as squids, have evolved a closed circulatory system and are able to move rapidly to catch prey. Closed circulatory systems are found in all vertebrates; however, there are significant differences in the structure of the heart and the circulation of blood between the different vertebrate groups due to adaptation during evolution and associated differences in anatomy. **Figure 27.10** illustrates the basic circulatory systems of some vertebrates: fish, amphibians, reptiles, and mammals.

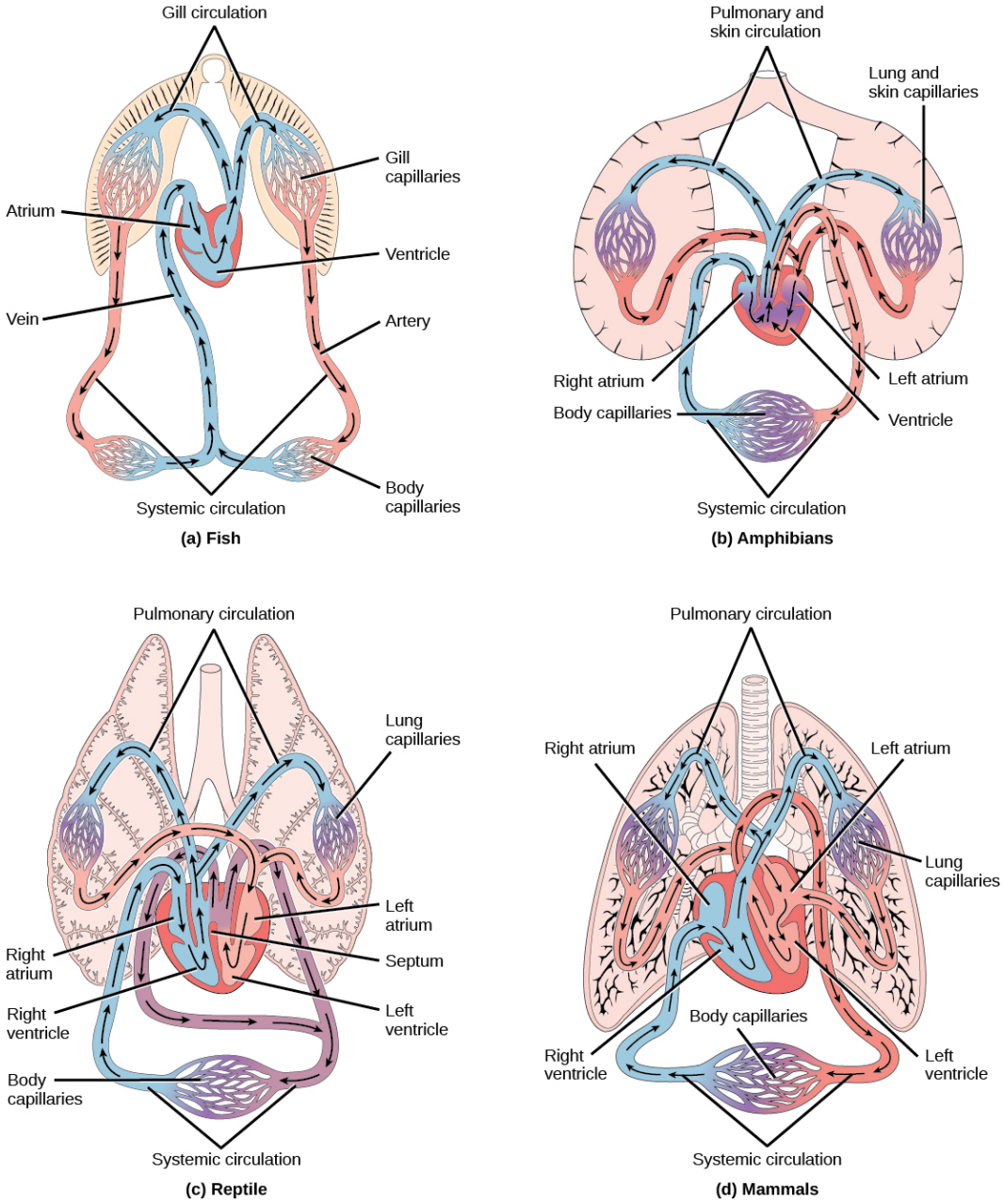


Figure 27.10 (a) Fish have the simplest circulatory systems of the vertebrates: blood flows unidirectionally from the two-chambered heart through the gills and then the rest of the body. (b) Amphibians have two circulatory routes: one for oxygenation of the blood through the lungs and skin, and the other to take oxygen to the rest of the body. The blood is pumped from a three-chambered heart with two atria and a single ventricle. (c) Reptiles also have two circulatory routes; however, blood is only oxygenated through the lungs. The heart is three chambered, but the ventricles are partially separated so some mixing of oxygenated and deoxygenated blood occurs except in crocodilians and birds. (d) Mammals and birds have the most efficient heart with four chambers that completely separate the oxygenated and deoxygenated blood; it pumps only oxygenated blood through the body and deoxygenated blood to the lungs.

As illustrated in **Figure 27.10a** Fish have a single circuit for blood flow and a two-chambered heart that has only a single atrium and a single ventricle. The atrium collects blood that has returned from the body, and the ventricle pumps the blood to the gills where gas exchange occurs and the blood is re-oxygenated; this is called the gill circulation. The blood then continues through the rest of the body before arriving back at the atrium; this is called the **systemic circulation**. This unidirectional flow of blood produces a gradient of oxygenated to deoxygenated blood around the fish's systemic circuit. The result is a limit in the amount of oxygen that can reach some of the organs and tissues of the body, reducing the overall metabolic capacity of fish.

In amphibians, reptiles, birds, and mammals, blood flow is directed in two circuits: one through the lungs and back to the heart, which is called the **pulmonary circulation**, and the other throughout the rest of the body and its organs including the brain (systemic circulation). In amphibians, gas exchange also occurs through the skin during pulmonary circulation and is referred to as pulmocutaneous circulation.

As shown in **Figure 27.10b**, amphibians have a three-chambered heart that has two atria and one ventricle rather than the two-chambered heart of fish. The two **atria** (superior heart chambers) receive blood from the two different circuits (the lungs and the systems), and then there is some mixing of the blood in the heart's **ventricle** (inferior heart chamber), which reduces the oxygen concentration in the blood pumped from the ventricle. The advantage to this arrangement is that high pressure in the vessels pushes blood to the lungs and body. The mixing is mitigated by a ridge within the ventricle that diverts oxygen-rich blood through the systemic circulatory system and deoxygenated blood to the pulmocutaneous circuit. For this reason, amphibians are often described as having double circulation.

Most reptiles also have a three-chambered heart similar to the amphibian heart that directs blood to the pulmonary and systemic circuits, as shown in **Figure 27.10c**. The ventricle is divided more effectively by a partial septum, which results in even less mixing of oxygenated and deoxygenated blood. Some reptiles (alligators and crocodiles) are the most primitive animals to exhibit a four-chambered heart. Crocodilians have a unique circulatory mechanism where the heart shunts blood from the lungs toward the stomach and other organs during long periods of submergence, for instance, while the animal waits for prey or stays underwater waiting for prey to rot. One adaptation includes two main arteries that leave the same part of the heart: one takes blood to the lungs and the other provides an alternate route to the stomach and other parts of the body. Two other adaptations include a hole in the heart between the two ventricles, called the foramen of Panizza, which allows blood to move from one side of the heart to the other, and specialized connective tissue that slows the blood flow to the lungs. Together these adaptations have made crocodiles and alligators one of the most successful and ancient animal groups on earth.

In mammals and birds, the heart is also divided into four chambers: two atria and two ventricles, as illustrated in **Figure 27.10d**. The oxygenated blood is completely separated from the deoxygenated blood, which improves the efficiency of double circulation and is probably required for the warm-blooded lifestyle of mammals and birds. The four-chambered heart of birds and mammals evolved independently from ancestors with a three-chambered heart. The independent evolution of the same or a similar biological trait is referred to as convergent evolution.

Components of Blood

Oxygen-binding proteins (hemoglobin, hemocyanin, etc.) are one of the main components of blood in all animals. The blood is more than those proteins, though. Blood is actually a term used to describe the liquid that moves through the vessels and includes **plasma** (the liquid portion, which contains water, proteins, salts, lipids, and glucose) and the cells (red and white cells) and cell fragments called **platelets**. Plasma is actually the major component of blood and contains the water, proteins, electrolytes, lipids, and glucose. The cells are responsible for carrying the gases (red cells) and immune the response (white). The platelets are responsible for blood clotting. In humans, cellular components make up approximately 45 percent of the blood and the liquid plasma 55 percent. Blood is 20 percent of a human's extracellular fluid and eight percent of the weight of an average human.

The Role of Blood in the Body

Blood, like the human blood illustrated in **Figure 27.11** is important for regulation of the body's systems and homeostasis. Blood helps maintain homeostasis by stabilizing pH, temperature, osmotic pressure, and by eliminating excess heat. Blood supports growth by distributing nutrients and hormones, and by removing waste. Blood plays a protective role by transporting clotting factors and platelets to prevent blood loss and transporting the disease-fighting agents or **white blood cells** to sites of infection.

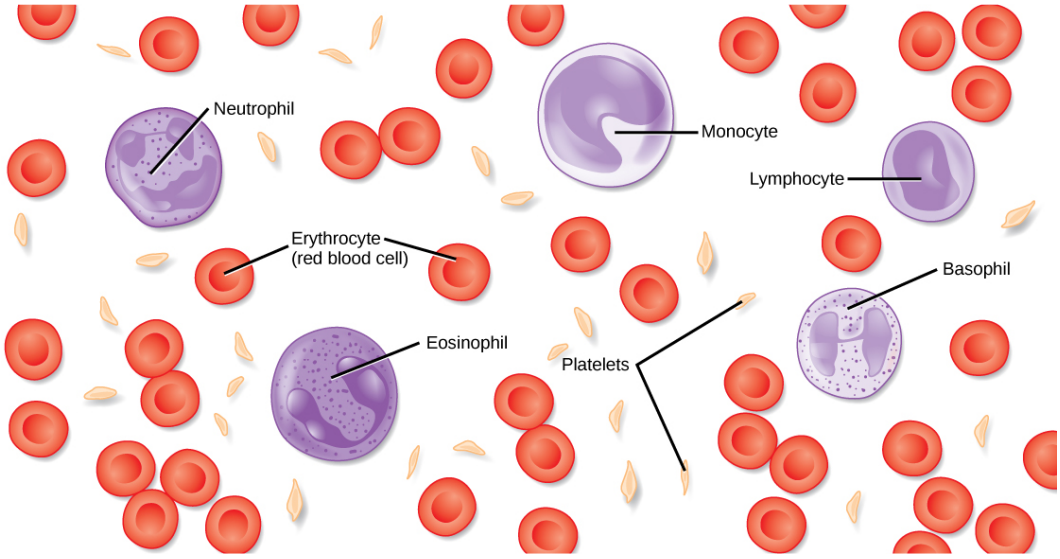


Figure 27.11 The cells and cellular components of human blood are shown. Red blood cells deliver oxygen to the cells and remove carbon dioxide. White blood cells—including neutrophils, monocytes, lymphocytes, eosinophils, and basophils—are involved in the immune response. Platelets form clots that prevent blood loss after injury.

Red Blood Cells

Red blood cells, or erythrocytes (erythro- = "red"; -cyte = "cell"), are specialized cells that circulate through the body delivering oxygen to cells; they are generated by division of stem cells in the bone marrow. In mammals, red blood cells are small biconcave cells that at maturity do not contain a nucleus or mitochondria and are only 7–8 µm in size. In birds and reptiles, erythrocytes have nuclie and mitochondria.

The red coloring of human blood comes from the iron-containing protein hemoglobin, illustrated in **Figure 27.12a**. The principal job of these proteins is to carry oxygen, but they also transports carbon dioxide as well. Hemoglobin is packed into human red blood cells at a rate of about 250 million molecules of hemoglobin per cell. Each hemoglobin molecule binds four oxygen molecules so that each red blood cell carries one billion molecules of oxygen. There are approximately 25 trillion red blood cells in the five liters of blood in the human body, which could carry up to 25 sextillion (25×10^{21}) molecules of oxygen in the body at any time. In mammals, the lack of organelles in erythrocytes leaves more room for the hemoglobin molecules, and the lack of mitochondria also prevents use of the oxygen for metabolic respiration.

Not all organisms use hemoglobin as the method of oxygen transport. Invertebrates that utilize hemolymph rather than blood use different pigments to bind to the oxygen. These pigments use copper or iron to the oxygen. Invertebrates have a variety of other respiratory pigments. Hemocyanin, a blue-green, copper-containing protein, illustrated in **Figure 27.12b** is found in mollusks, crustaceans, and some of the arthropods. Chlorocruorin, a green-colored, iron-containing pigment is found in four families of polychaete tubeworms. Hemerythrin, a red, iron-containing protein is found in some polychaete worms and annelids and is illustrated in **Figure 27.12c**. Despite the name, hemerythrin does not contain a heme group and its oxygen-carrying capacity is poor compared to hemoglobin.

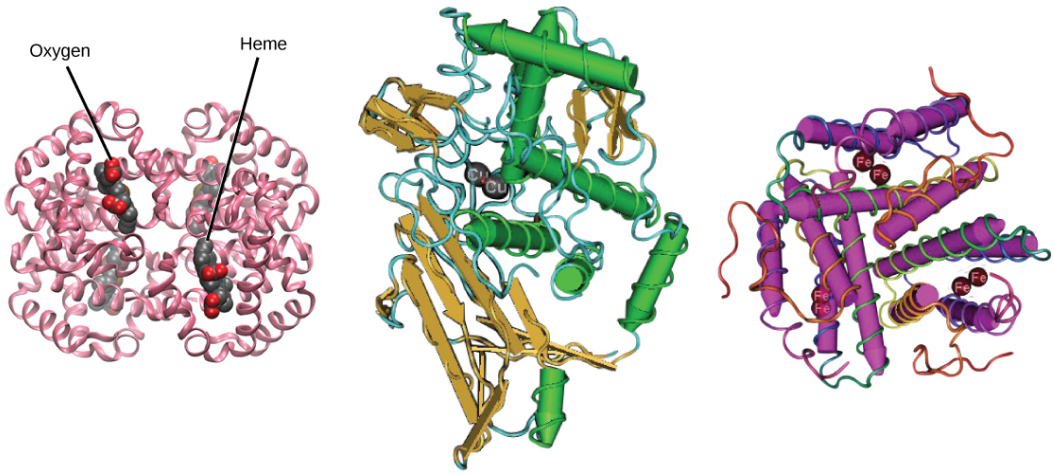


Figure 27.12 In most vertebrates, (a) hemoglobin delivers oxygen to the body and removes some carbon dioxide. Hemoglobin is composed of four protein subunits, two alpha chains and two beta chains, and a heme group that has iron associated with it. The iron reversibly associates with oxygen, and in so doing is oxidized from Fe^{2+} to Fe^{3+}. In most mollusks and some arthropods, (b) hemocyanin delivers oxygen. Unlike hemoglobin, hemolymph is not carried in blood cells, but floats free in the hemolymph. Copper instead of iron binds the oxygen, giving the hemolymph a blue-green color. In annelids, such as the earthworm, and some other invertebrates, (c) hemerythrin carries oxygen. Like hemoglobin, hemerythrin is carried in blood cells and has iron associated with it, but despite its name, hemerythrin does not contain heme.

The small size and large surface area of red blood cells allows for rapid diffusion of oxygen and carbon dioxide across the plasma membrane. In the lungs, carbon dioxide is released and oxygen is taken in by the blood. In the tissues, oxygen is released from the blood and carbon dioxide is bound for transport back to the lungs.

White Blood Cells

White blood cells, also called leukocytes (leuko = white), make up approximately one percent (by volume) of the cells in blood. The role of white blood cells is very different than that of red blood cells, as you have learned previously. They are primarily involved in the immune response to identify and target pathogens, such as invading bacteria, viruses, and other foreign organisms. White blood cells are formed continually; some only live for hours or days, but some live for years.

The morphology of white blood cells differs significantly from red blood cells. They have nuclei and do not contain hemoglobin. The different types of white blood cells are identified by their microscopic appearance after histologic staining, and each has a different specialized function. The two main groups, both illustrated in **Figure 27.11** are the granulocytes, which include the neutrophils, eosinophils, and basophils, and the agranulocytes, which include the monocytes and lymphocytes.

Granulocytes contain granules in their cytoplasm; the agranulocytes are so named because of the lack of granules in their cytoplasm. Some leukocytes become macrophages that either stay at the same site or move through the blood stream and gather at sites of infection or inflammation where they are attracted by chemical signals from foreign particles and damaged cells. Lymphocytes are the primary cells of the immune system and include B cells, T cells, and natural killer cells. B cells destroy bacteria and inactivate their toxins. They also produce antibodies. T cells attack viruses, fungi, some bacteria, transplanted cells, and cancer cells. T cells attack viruses by releasing toxins that kill the viruses. Natural killer cells attack a variety of infectious microbes and certain tumor cells.

Platelets and Coagulation Factors

Blood must clot to heal wounds and prevent excess blood loss. Small cell fragments called platelets (thrombocytes) are attracted to the wound site where they adhere by extending many projections and releasing their contents. These contents activate other platelets and also interact with other coagulation factors, which convert fibrinogen, a water-soluble protein present in blood serum into fibrin (a non-water soluble protein), causing the blood to clot. Many of the clotting factors require vitamin K to work, and vitamin K deficiency can lead to problems with blood clotting. Many platelets converge and stick together at the wound site forming a platelet plug (also called a fibrin clot), as illustrated in **Figure 27.13b**. The plug or clot lasts for a number of days and stops the loss of blood. Platelets are formed from the disintegration of larger cells called megakaryocytes, like that shown in **Figure 27.13a**. For each megakaryocyte, 2000–3000 platelets are formed with 150,000 to 400,000 platelets present in each cubic millimeter of blood. Each platelet is disc shaped and 2–4 μm in diameter.

They contain many small vesicles but do not contain a nucleus.

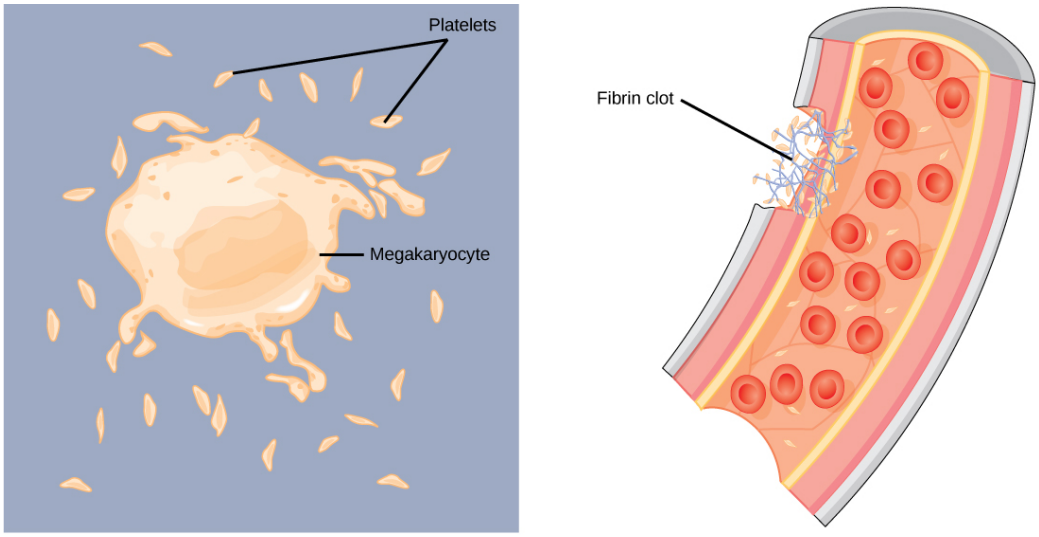


Figure 27.13 (a) Platelets are formed from large cells called megakaryocytes. The megakaryocyte breaks up into thousands of fragments that become platelets. (b) Platelets are required for clotting of the blood. The platelets collect at a wound site in conjunction with other clotting factors, such as fibrinogen, to form a fibrin clot that prevents blood loss and allows the wound to heal.

Plasma and Serum

The liquid component of blood is called plasma, and it can separated from the blood cells by spinning or centrifuging the blood at high rotations (3000 rpm or higher). The blood cells and platelets are separated by centrifugal forces to the bottom of a specimen tube. The upper liquid layer, the plasma, consists of 90 percent water along with various substances required for maintaining the body's pH, osmotic load, and for protecting the body. The plasma also contains the coagulation factors and antibodies.

The plasma component of blood without the coagulation factors is called the **serum**. Serum is similar to interstitial fluid in which the correct composition of key ions acting as electrolytes is essential for normal functioning of muscles and nerves. Other components in the serum include proteins that assist with maintaining pH and osmotic balance while giving viscosity to the blood. The serum also contains antibodies, specialized proteins that are important for defense against viruses and bacteria. Lipids, including cholesterol, are also transported in the serum, along with various other substances including nutrients, hormones, metabolic waste, plus external substances, such as, drugs, viruses, and bacteria.

evolution CONNECTION

Blood Types Related to Proteins on the Surface of the Red Blood Cells

Red blood cells are coated in antigens made of glycolipids and glycoproteins. The composition of these molecules is determined by genetics, which have evolved over time. In humans, the different surface antigens are grouped into 24 different blood groups with more than 100 different antigens on each red blood cell. The two most well known blood groups are the ABO, shown in **Figure 27.14**, and Rh systems. The surface antigens in the ABO blood group are glycolipids, called antigen A and antigen B. People with blood type A have antigen A, those with blood type B have antigen B, those with blood type AB have both antigens, and people with blood type O have neither antigen. Antibodies are found in the blood plasma can react with the A or B antigens; individuals with type A blood have antibodies to thye B blood cells. When type A and type B blood are combined, agglutination (clumping) of the blood occurs because of antibodies in the plasma that bind with the opposing antigen; this causes clots that coagulate in the kidney causing kidney failure and death. Type O blood has neither A or B antigens, and therefore, type O blood can be given to all blood types. Type O negative blood is the universal donor. Type AB positive blood is the universal acceptor because it has both A and B antigen.

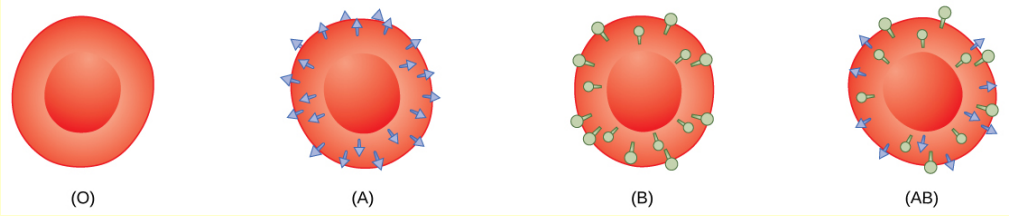


Figure 27.14 Human red blood cells may have either type A or B glycoproteins on their surface, both glycoproteins combined (AB), or neither (O). The glycoproteins serve as antigens and can elicit an immune response in a person who receives a transfusion containing unfamiliar antigens. Type O blood, which has no A or B antigens, does not elicit an immune response when injected into a person of any blood type. Thus, O is considered the universal donor. Persons with type AB blood can accept blood from any blood type, and type AB is considered the universal acceptor.

Mammalian Heart and Blood Vessels

The heart is a complex muscle that pumps blood through the three divisions of the circulatory system: the coronary (vessels that serve the heart), pulmonary (heart and lungs), and systemic (systems of the body), as shown in **Figure 27.15**. Coronary circulation is intrinsic to the heart, and takes blood directly from the main artery (aorta) coming from the heart to provide oxygen for the hard-working heart muscle. For pulmonary and systemic circulation, the heart has to pump blood to the lungs or the rest of the body, respectively. In vertebrates, the lungs are relatively close to the heart in the thoracic cavity. The shorter distance to pump means that the heart muscle wall on the right side of the heart is not as thick as the left side which must have enough pressure to pump blood all the way to your big toe.

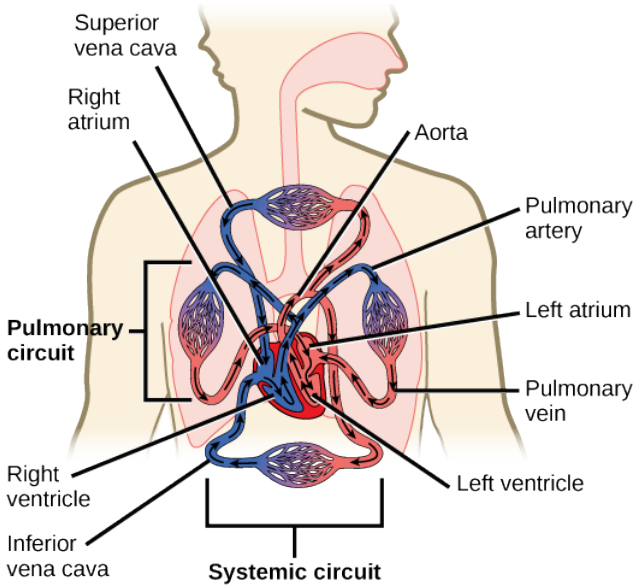


Figure 27.15 The mammalian circulatory system is divided into three circuits: the systemic circuit, the pulmonary circuit, and the coronary circuit. Blood is pumped from veins of the systemic circuit into the right atrium of the heart, then into the right ventricle. Blood then enters the pulmonary circuit, and is oxygenated by the lungs. From the pulmonary circuit, blood re-enters the heart through the left atrium. From the left ventricle, blood re-enters the systemic circuit through the aorta and is distributed to the rest of the body. The coronary circuit, which provides blood to the heart, is not shown.

Structure of the Heart

The heart muscle is asymmetrical as a result of the distance blood must travel in the pulmonary and systemic circuits. Since the right side of the heart sends blood to the pulmonary circuit it is smaller than the left side which must send blood out to the whole body in the systemic circuit, as shown in **Figure 27.15**. 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 atria are the chambers that receive blood from the circulation, and the ventricles are the chambers that pump blood into the circulation. The right atrium receives deoxygenated blood from the **superior vena cava**, which drains blood from the jugular vein that comes from the brain and from the veins that come from the arms, as well as from the **inferior vena cava** which drains blood from the veins that come from the lower organs and the legs. In addition, the right atrium receives blood from the coronary sinus which drains deoxygenated blood from the heart itself. This deoxygenated blood then passes to the right ventricle through the atrioventricular valve or the tricuspid valve. After it is filled, the right ventricle pumps the blood through the pulmonary arteries, by-passing the semilunar valve (or pulmonic valve) to the lungs for re-oxygenation. After blood passes through the pulmonary arteries, the right semilunar valves close preventing the blood from flowing backwards into the right ventricle. The left atrium then receives the oxygen-rich blood from the lungs via the pulmonary veins. This blood passes through the bicuspid valve or mitral valve (the atrioventricular valve on the left side of the heart) to the left ventricle where the blood is pumped out through **aorta**, the major artery of the body, taking oxygenated blood to the organs and muscles of the body. Once blood is pumped out of the left ventricle and into the aorta, the aortic semilunar valve (or aortic valve) closes preventing blood from flowing backward into the left ventricle. This pattern of pumping is referred to as double circulation and is found in all mammals.

The heart has its own blood vessels that supply the heart muscle with blood. The coronary arteries branch from the aorta and surround the outer surface of the heart like a crown. They diverge into capillaries where the heart muscle is supplied with oxygen before converging again into the coronary veins to take the deoxygenated blood back to the right atrium where the blood will be re-oxygenated through the pulmonary circuit. The heart muscle will die without a steady supply of blood. **Atherosclerosis** is the blockage of an artery by the buildup of fatty plaques. Because of the size (narrow) of the coronary arteries and their function in serving the heart itself, atherosclerosis can be deadly in these arteries. The slowdown of blood flow and subsequent oxygen deprivation that results from atherosclerosis causes severe pain, known as angina, and complete blockage of the arteries will cause myocardial infarction: the death of cardiac muscle tissue, commonly known as a heart attack.

Arteries, Veins, and Capillaries

The blood from the heart is carried through the body by a complex network of blood vessels (**Figure 27.16**). **Arteries** take blood away from the heart. The main artery is the aorta that branches into major arteries that take blood to different limbs

and organs. These major arteries include the carotid artery that takes blood to the brain, the brachial arteries that take blood to the arms, and the thoracic artery that takes blood to the thorax and then into the hepatic, renal, and gastric arteries for the liver, kidney, and stomach, respectively. The iliac artery takes blood to the lower limbs. 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.

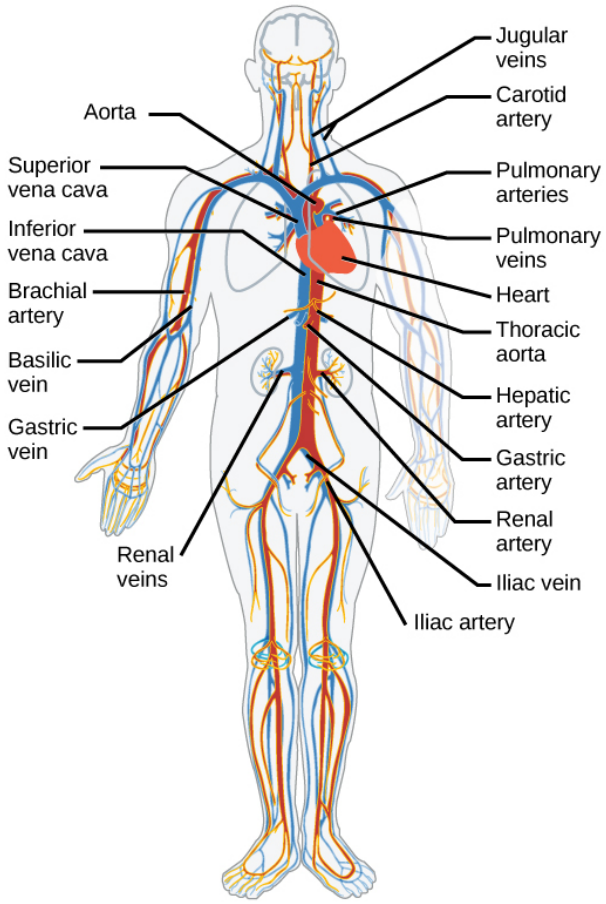


Figure 27.16 The major human arteries and veins are shown. (credit: modification of work by Mariana Ruiz Villareal)

Arterioles diverge into capillary beds. **Capillary beds** contain a large number (10 to 100) of **capillaries** that branch among the cells and tissues of the body. Capillaries are narrow-diameter tubes that can fit red blood cells through in single file and are the sites for the exchange of nutrients, waste, and oxygen with tissues at the cellular level. Fluid also crosses into the interstitial space from the capillaries. The capillaries converge again into **venules** that connect to minor veins that finally connect to major veins that take blood high in carbon dioxide back to the heart. **Veins** are blood vessels that bring blood back to the heart. The major veins drain blood from the same organs and limbs that the major arteries supply. Fluid is also brought back to the heart via the lymphatic system.

27.3 | Innate Immunity

Introduction

> “The organism possesses certain contrivances by means of which the immunity reaction, so easily produced by all kinds of cells, is prevented from acting against the organism's own elements and so giving rise to autotoxins... These contrivances are naturally of the

highest importance for the existence of the individual.”

Paul Ehrlich, German immunologist and pharmacologist, "On Hemolysins", 1901

Although other organisms have immune responses of various types, the vertebrate immune system is the most sophisticated. The vertebrate immune system is a complex multilayered system for defending against external and internal threats to the integrity of the body. The system can be divided into two types of defense systems: the **innate immune system**, which is nonspecific toward a particular kind of pathogen, and the **adaptive immune system**, which is specific (**Figure 27.17**). Innate immunity relies on physical and chemical barriers that work on all pathogens, sometimes called the first line of defense. The second line of defense of the innate system includes chemical signals that produce inflammation and fever responses as well as mobilizing protective cells and other chemical defenses. The adaptive immune system mounts a highly specific response to substances and organisms that do not belong in the body. The adaptive system takes longer to respond and has a memory system that allows it to respond with greater intensity should the body re-encounter a pathogen, even years later. That memory is the basis for the long-term effectiveness of many vaccines.

Vertebrate Immunity		
Innate Immune System		Adaptive Immune System
Physical Barriers	Internal Defenses	
• Skin, hair, cilia • Mucus membranes • Mucus and chemical secretions • Digestive enzymes in mouth • Stomach acid	• Inflammatory response • Complement proteins • Phagocytic cells • Natural killer (NK) cells	• Antibodies and the humoral immune response • Cell-mediated immune response • Memory response

Figure 27.17 There are two main parts to the vertebrate immune system. The innate immune system, which is made up of physical barriers and internal defenses, responds to all pathogens. The adaptive immune system is highly specific.

External and Chemical Barriers

The body has significant physical barriers to potential pathogens. The skin contains the protein keratin, which resists physical entry into cells. Other body surfaces, particularly those associated with body openings, are protected by the mucous membranes. The sticky mucus provides a physical trap for pathogens, preventing their movement deeper into the body. Some openings of the body, such as the nose and ears, are protected by hairs that catch pathogens, and the mucous membranes of the upper respiratory tract have cilia that constantly sweep pathogens trapped in the mucus coat up to the mouth. The skin and mucous membranes also create a chemical environment that is hostile to many microorganisms. The surface of the skin is acidic, which prevents bacterial growth. Saliva, mucus, and the tears of the eye contain an enzyme that breaks down bacterial cell walls. The stomach secretions create a highly acidic environment, which kills many pathogens entering the digestive system.

Finally, the surface of the body and the lower digestive system have a community of commensal microorganisms such as bacteria, archaea, and fungi (the microbiome) that coexist without harming the body. There is evidence that these organisms are highly beneficial to their host, combating disease-causing organisms and outcompeting them for nutritional resources provided by the host body. Despite these defenses, pathogens may enter the body through skin abrasions or punctures, or by collecting on mucosal surfaces in large numbers that overcome the protections of mucus or cilia.

Internal Defenses

If pathogens defeat these defenses and enter the body, the innate immune system responds with a variety of internal defenses. These include the inflammatory response, phagocytosis, natural killer cells, and the complement system. White blood cells in the blood and lymph recognize pathogens as foreign to the body. A **white blood cell** is larger than a red blood cell, is nucleated, and is typically able to move using amoeboid locomotion. Because they can move on their own, white blood cells can leave the blood to go to infected tissues. For example, a monocyte is a type of white blood cell that circulates in the blood and lymph and develops into a macrophage after it moves into infected tissue. A macrophage is a large phagocytic cell that engulfs and devours foreign particles and pathogens.

Once a pathogen is recognized as foreign and devoured by a macrophage, chemicals called cytokines are released. A cytokine is a chemical messenger that regulates cell differentiation (form and function), proliferation (production), and gene

expression to produce a variety of immune responses. Approximately 40 types of cytokines exist in humans. In addition to being released from white blood cells after pathogen recognition, cytokines are also released by the infected cells and bind to nearby uninfected cells, inducing those cells to release cytokines. This positive feedback loop results in a burst of cytokine production.

One class of early-acting cytokines is the interferons, which are released by infected cells as a warning to nearby uninfected cells. An interferon is a small protein that signals a viral infection to other cells. The interferons stimulate uninfected cells to produce compounds that interfere with viral replication. Interferons also activate macrophages and other cells.

The Inflammatory Response and Phagocytosis

The first cytokines to be produced encourage **inflammation**, a localized redness, swelling, heat, and pain. Inflammation is a response to physical trauma, such as a cut or a blow, chemical irritation, and infection by pathogens (viruses, bacteria, or fungi). The chemical signals that trigger an inflammatory response enter the extracellular fluid and cause capillaries to dilate (expand) and capillary walls to become more permeable, or leaky. The serum and other compounds leaking from capillaries cause swelling of the area, which in turn causes pain. Various kinds of white blood cells are attracted by the cytokines released at the area of inflammation. The types of white blood cells that arrive at an inflamed site depend on the nature of the injury or infecting pathogen. For example, a neutrophil is an early arriving white blood cell that engulfs and digests pathogens. Neutrophils are the most abundant white blood cells of the immune system (**Figure 27.18**). Macrophages follow neutrophils and take over the phagocytosis function and are involved in cleaning up cell debris and pathogens.

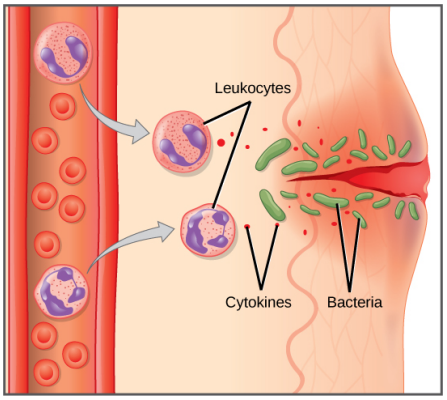


Figure 27.18 White blood cells (leukocytes) release chemicals to stimulate the inflammatory response following a cut in the skin.

Cytokines also send feedback to cells of the nervous system to bring about the overall symptoms of feeling sick, which include lethargy, muscle pain, and nausea. Cytokines can thus increase the core body temperature, causing a fever. The elevated temperatures of a fever inhibit the growth of pathogens and speed up cellular repair processes. For these reasons, suppression of fevers should be limited to those that are dangerously high.

Natural Killer Cells

A **lymphocyte** is a white blood cell that contains a large nucleus (**Figure 27.19**). Most lymphocytes are associated with the adaptive immune response, but a class of lymphocytes known as natural killer cells are part of the innate immune system. Unlike other white blood cells that attack invading bacteria or fungi, **natural killer (NK) cell** is a lymphocyte. Natural killer cells kill body cells that are infected with viruses (or cancerous cells). NK cells identify intracellular infections, especially from viruses, and attack the infected cells, destroying them so that they cannot release more viruses.

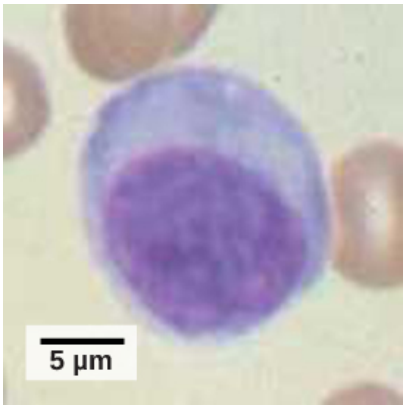


Figure 27.19 Lymphocytes, such as NK cells, are characterized by their large nuclei that actively absorb Wright stain and therefore appear dark colored under a microscope. (credit: scale-bar data from Matt Russell)

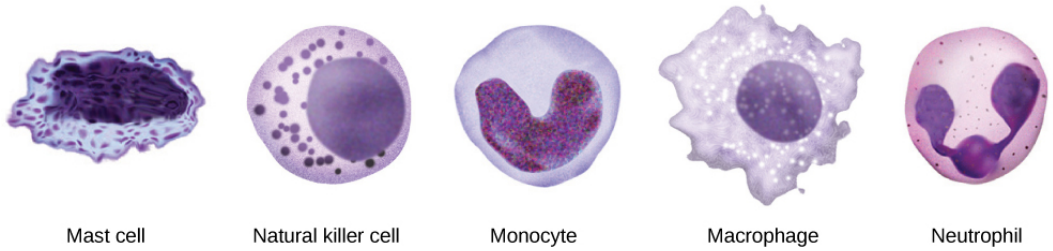


Figure 27.20 Cells involved in the innate immune response include mast cells, natural killer cells, and white blood cells, such as monocytes, macrophages and neutrophils.

Complement

An array of approximately 20 types of proteins, called a complement system, is also activated by infection or the activity of the cells of the adaptive immune system and functions to destroy extracellular pathogens. Liver cells and macrophages synthesize inactive forms of complement proteins continuously; these proteins are abundant in the blood serum and are capable of responding immediately to infecting microorganisms. The complement system is so named because it is complementary to the innate and adaptive immune system. Complement proteins bind to the surfaces of microorganisms and are particularly attracted to pathogens that are already tagged by the adaptive immune system. This “tagging” involves the attachment of specific proteins called antibodies (discussed in detail later) to the pathogen. When they attach, the antibodies change shape providing a binding site for one of the complement proteins. After the first few complement proteins bind, a cascade of binding in a specific sequence of proteins follows in which the pathogen rapidly becomes coated in complement proteins.

Complement proteins perform several functions, one of which is to serve as a marker to indicate the presence of a pathogen to phagocytic cells and enhance engulfment. Certain complement proteins can combine to open pores in microbial plasma membranes, causing ion leakage and lysis of the microbial cells.

27.4 | Adaptive Immunity

Introduction

> “An immune system of enormous complexity is present in all vertebrate animals. When we place a population of lymphocytes from such an animal in appropriate tissue culture fluid, and when we add an antigen, the lymphocytes will produce specific antibody molecules, in the absense of any nerve cells. I find it astonishing

> that the immune system embodies a degree of complexity which suggests some more or less superficial though striking analogies with human language, and that this cognitive system has evolved and functions without assistance of the brain.”

Niels K. Jerne, Danish immunologist, "The Generative Grammar of the Immune System", Nobel Lecture, 1984

The adaptive, or acquired, immune response takes days or even weeks to become established—much longer than the innate response; however, adaptive immunity is more specific to an invading pathogen. **Adaptive immunity** is an immunity that occurs after exposure to an antigen either from a pathogen or a vaccination. An **antigen** is a molecule that stimulates a response in the immune system. This part of the immune system is activated when the innate immune response is insufficient to control an infection. In fact, without information from the innate immune system, the adaptive response could not be mobilized. There are two types of adaptive responses: the **cell-mediated immune response**, which is controlled by activated **T cells**, and the **humoral immune response**, which is controlled by activated **B cells** and antibodies. Activated T and B cells, which specifically bind to molecules from the invading pathogen, attack the pathogen specifically. These cells can kill pathogens directly, or they can secrete antibodies that enhance the phagocytosis of pathogens and disrupt the infection. Adaptive immunity also involves a memory to give the host long-term protection from reinfection with the same type of pathogen; on reexposure, this host memory will facilitate a rapid and powerful response.

B and T Cells

Lymphocytes, which are white blood cells, are formed with other blood cells in the red bone marrow. The two types of lymphocytes of the adaptive immune response are B and T cells (Figure 27.21). Whether an immature lymphocyte becomes a B cell or T cell depends on where in the body it matures. The B cells remain in the bone marrow to mature (hence the name “B” for “bone marrow”), while T cells migrate to the thymus, where they mature (hence the name “T” for “thymus”).

Maturation of a B or T cell involves becoming immunocompetent, meaning that it can recognize, by binding, a specific molecule or antigen (discussed below). During the maturation process, B and T cells that bind too strongly to the body’s own cells are eliminated in order to minimize an immune response against the body’s own tissues. Those cells that react weakly or not at all to the body’s own cells, but have highly specific receptors on their cell surfaces that allow them to recognize a foreign molecule, or antigen, remain. This process occurs during fetal development and continues throughout life. The specificity of this receptor is determined by the genetics of the individual and is present before a foreign molecule is introduced to the body or encountered. Thus, it is genetics and not experience that initially provides a vast array of cells, each capable of binding to a different specific foreign molecule. Once they are immunocompetent, the T and B cells will migrate to the spleen and lymph nodes where they will remain until they are called on during an infection. B cells are involved in the humoral immune response, which targets pathogens found in blood and lymph, and T cells are involved in the cell-mediated immune response, which targets infected body cells.

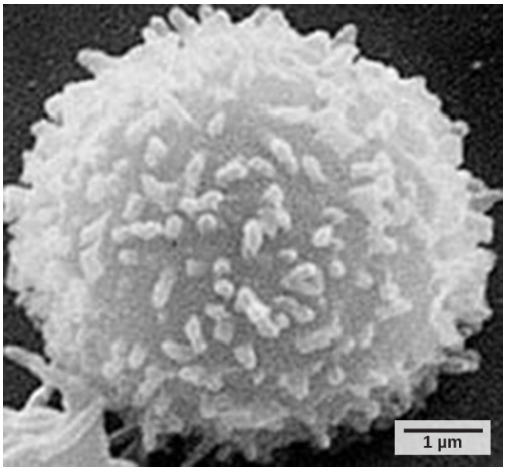


Figure 27.21 This scanning electron micrograph shows a T lymphocyte. T and B cells are indistinguishable by light microscopy but can be differentiated experimentally by probing their surface receptors. (credit: modification of work by NCI; scale-bar data from Matt Russell)

Humoral Immune Response

As mentioned, an antigen is a molecule that stimulates a response in the immune system. B cells participate in a chemical response to new antigens by producing specific antibodies that circulate throughout the body and bind with the antigen whenever it is encountered. This is known as the **humoral immune response**, because the active molecule is secreted into the body fluids, or "humours". As discussed, during maturation of B cells, a set of highly specific B cells are produced that have many antigen receptor molecules in their membrane (Figure 27.22).

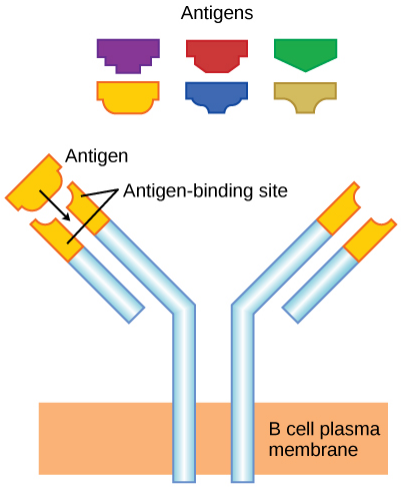


Figure 27.22 B cell receptors are embedded in the membranes of B cells and bind a variety of antigens through their variable regions.

Each B cell has only one kind of antigen receptor, which makes every B cell different. Once the B cells mature in the bone marrow, they migrate to lymph nodes or other lymphatic organs. When a B cell encounters the antigen that binds to its receptor, the antigen molecule is brought into the cell by endocytosis. It is digested in the lysosomes, and fragments of the foreign molecule are then displayed on the surface of the B cell. These displayed molecules can activate other cells as part of the immune response. When this process is complete, the B cell is sensitized. In most cases, the sensitized B cell must then encounter a specific kind of T cell, called a helper T cell, before it is activated. The helper T cell must already have been activated through an encounter with the antigen (discussed below).

The helper T cell binds to the displayed antigen fragments on the B cell, and is activated to release cytokines that induce the B cell to divide rapidly. This generates thousands of identical (clonal) B cells. These daughter cells have two possible fates: they can become either plasma cells or memory B cells. The **plasma cells** produce and secrete large quantities of antibody molecules, up to 100 million molecules per hour, . An **antibody**, also known as an immunoglobulin (Ig), is a protein that is produced by plasma cells after stimulation by an antigen. Antibodies are the agents of humoral immunity. Antibodies occur in the blood, in gastric and mucus secretions, and in breast milk. Antibodies in these bodily fluids can bind pathogens and mark them for destruction by phagocytes before they can infect cells. The **memory B cells** become quiescent, and only become reactivated after another later encounter with the antigen. This can be caused by a reinfection by the same bacteria or virus, and activation of the memory cells again results in a new population of antibody-producing plasma cells to fight the re-infection.

These antibodies circulate in the blood stream and lymphatic system and bind with the antigen whenever it is encountered. The binding can fight infection in several ways. Antibodies can bind to viruses or bacteria and interfere with the chemical interactions required for them to infect or bind to other cells. The antibodies may create bridges between different particles containing antigenic sites clumping them all together and preventing their proper functioning. The antigen-antibody complex stimulates the complement system described previously, destroying the cell bearing the antigen. Phagocytic cells, such as those already described, are attracted by the antigen-antibody complexes, and phagocytosis is enhanced when the complexes are present. Finally, antibodies stimulate inflammation, and their presence in mucus and on the skin prevents pathogen attack.

Antibodies coat extracellular pathogens and neutralize them by blocking key sites on the pathogen that enhance their infectivity (such as receptors that “dock” pathogens on host cells) (Figure 27.23). Antibody neutralization can prevent pathogens from entering and infecting host cells. The neutralized antibody-coated pathogens can then be filtered by the spleen and destroyed.

Antibodies also mark pathogens for destruction by phagocytic cells, such as macrophages or neutrophils, in a process called

opsonization. In a process called complement fixation, some antibodies provide a place for complement proteins to bind. The combination of antibodies and complement promotes rapid clearing of pathogens.

The production of antibodies by plasma cells in response to an antigen is called active immunity and describes the host's active immune response to an infection or to a vaccination. There is also a passive immune response where antibodies come from an outside source, instead of the individual's own plasma cells, and are introduced into the host. For example, antibodies circulating in a pregnant woman's body move across the placenta into the developing fetus. The child benefits from the presence of these antibodies for up to several months after birth. In addition, a passive immune response is possible by injecting antibodies into an individual in the form of an antivenom to a snake-bite toxin or antibodies in blood serum to help fight a hepatitis infection. This gives immediate protection since the body does not need the time required to mount its own response.

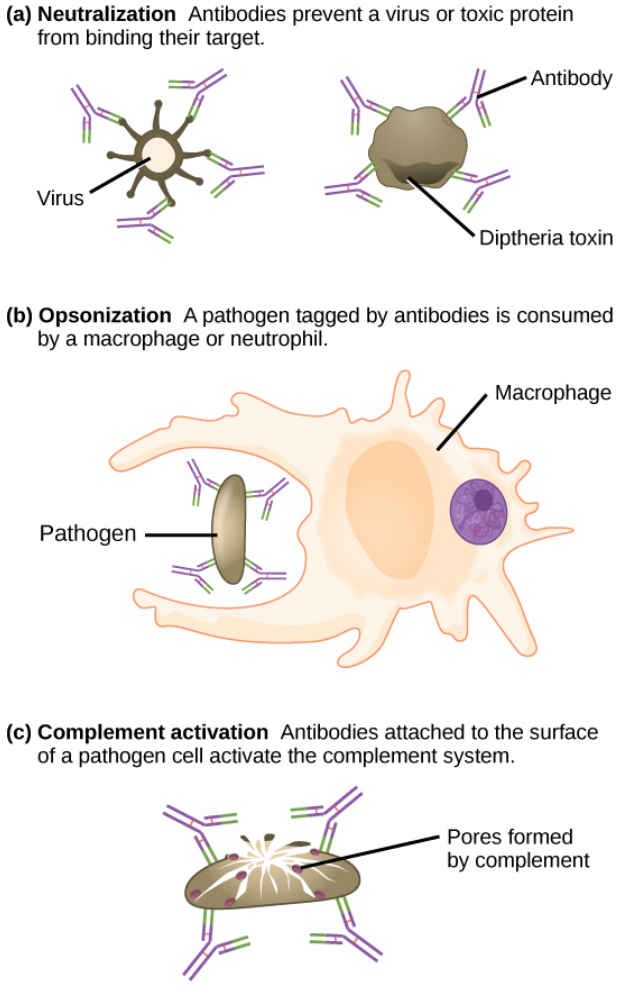


Figure 27.23 Antibodies may inhibit infection by (a) preventing the antigen from binding its target, (b) tagging a pathogen for destruction by macrophages or neutrophils, or (c) activating the complement cascade.

Cell-Mediated Immunity

Unlike B cells, T lymphocytes are unable to recognize pathogens without assistance. Instead, dendritic cells and macrophages first engulf and digest pathogens into hundreds or thousands of antigens. Then, an **antigen-presenting cell (APC)** detects, engulfs, and informs the adaptive immune response about an infection. When a pathogen is detected, these APCs will engulf and break it down through phagocytosis. Antigen fragments will then be transported to the surface of the APC, where they will serve as an indicator to other immune cells. A **dendritic cell** is an immune cell that mops up antigenic materials in its surroundings and presents them on its surface. Dendritic cells are located in the skin, the linings of the nose, lungs, stomach, and intestines. These positions are ideal locations to encounter invading pathogens. Once they are activated by pathogens and mature to become APCs they migrate to the spleen or a lymph node. Macrophages also function as APCs. In all cases the foreign antigen is digested inside the cell, and fragments of the antigen are then displayed on the surface of the APC.

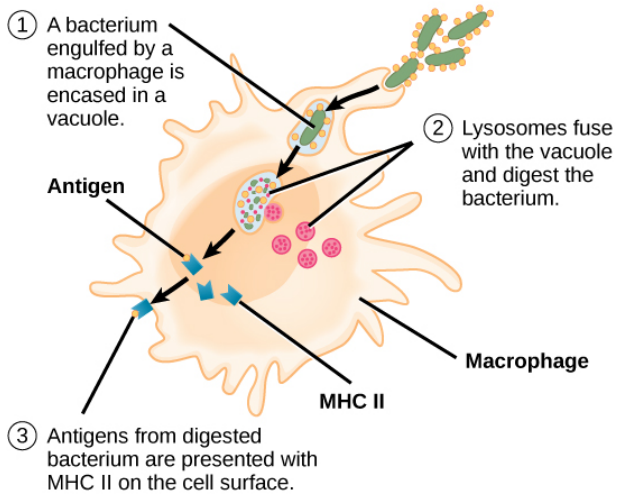


Figure 27.24 An antigen-presenting cell (APC), such as a macrophage, engulfs a foreign antigen, partially digests it in a lysosome, and then displays it at the cell surface. Lymphocytes of the adaptive immune response must interact with these displayed fragments, bound to a specific protein on the APC cell surface, in order to mature into functional immune cells.

T cells have many functions. Some respond to APCs of the innate immune system and indirectly induce immune responses by releasing cytokines. Others stimulate B cells to start the humoral response as described previously. Another type of T cell detects APC signals and directly kills the infected cells, while some are involved in suppressing inappropriate immune reactions to harmless or "self" antigens.

There are two main types of T cells: helper T lymphocytes (T_H) and the cytotoxic T lymphocytes (T_C). The T_H lymphocytes function indirectly to tell other immune cells about potential pathogens. T_H lymphocytes recognize specific antigens presented by the MHC class II complexes of APCs. There are two populations of T_H cells: T_H1 and T_H2. T_H1 cells secrete cytokines to enhance the activities of macrophages and other T cells. T_H2 cells stimulate naïve B cells to secrete antibodies. Whether a T_H1 or a T_H2 immune response develops depends on the specific types of cytokines secreted by cells of the innate immune system, which in turn depends on the nature of the invading pathogen.

Cytotoxic T cells (T_C) are the key component of the cell-mediated part of the adaptive immune system and attack and destroy infected cells. T_C cells are particularly important in protecting against viral infections; this is because viruses replicate within cells where they are shielded from extracellular contact with circulating antibodies. Once activated, the T_C creates a large clone of cells with one specific set of cell-surface receptors, as in the case with proliferation of activated B cells. As with B cells, the clone includes active T_C cells and inactive memory T_C cells. The resulting active T_C cells then identify infected host cells. Because of the time required to generate a population of clonal T and B cells, there is a delay in the adaptive immune response compared to the innate immune response.

T_C cells attempt to identify and destroy infected cells before the pathogen can replicate and escape, thereby halting the progression of intracellular infections. T_C cells also support NK lymphocytes to destroy early cancers. Cytokines secreted by the T_H1 response that stimulates macrophages also stimulate T_C cells and enhance their ability to identify and destroy infected cells and tumors. A summary of how the humoral and cell-mediated immune responses are activated appears in **Figure 27.25**.

B plasma cells and T_C cells are collectively called effector cells because they are involved in "effecting" (bringing about) the immune response of killing pathogens and infected host cells.

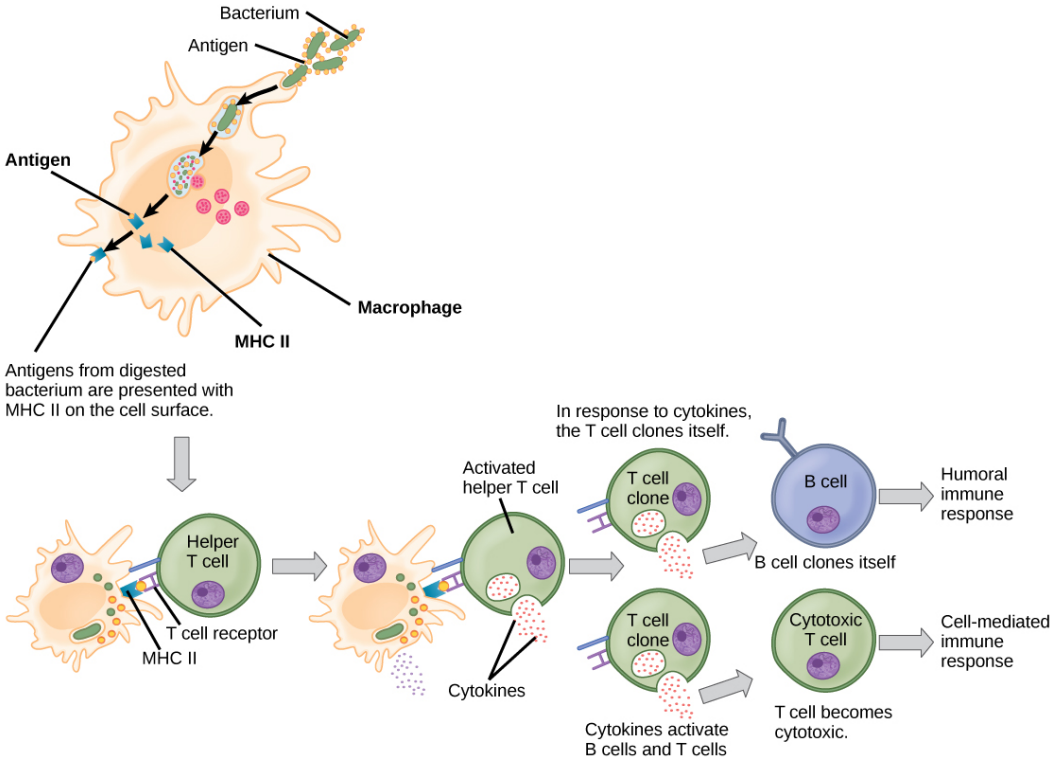


Figure 27.25 A helper T cell becomes activated by binding to an antigen presented by an APC via the MHCII receptor, causing it to release cytokines. Depending on the cytokines released, this activates either the humoral or the cell-mediated immune response.

Immunological Memory

The adaptive immune system has a memory component that allows for a rapid and large response upon reinvasion of the same pathogen. During the adaptive immune response to a pathogen that has not been encountered before, known as the primary immune response, plasma cells secreting antibodies and differentiated T cells increase, then plateau over time. As B and T cells mature into effector cells, a subset of the naïve populations differentiates into B and T memory cells with the same antigen specificities (**Figure 27.26**). A **memory cell** is an antigen-specific B or T lymphocyte that does not differentiate into an effector cell during the primary immune response, but that can immediately become an effector cell on reexposure to the same pathogen. As the infection is cleared and pathogenic stimuli subside, the effectors are no longer needed and they undergo apoptosis. In contrast, the memory cells persist in the circulation.

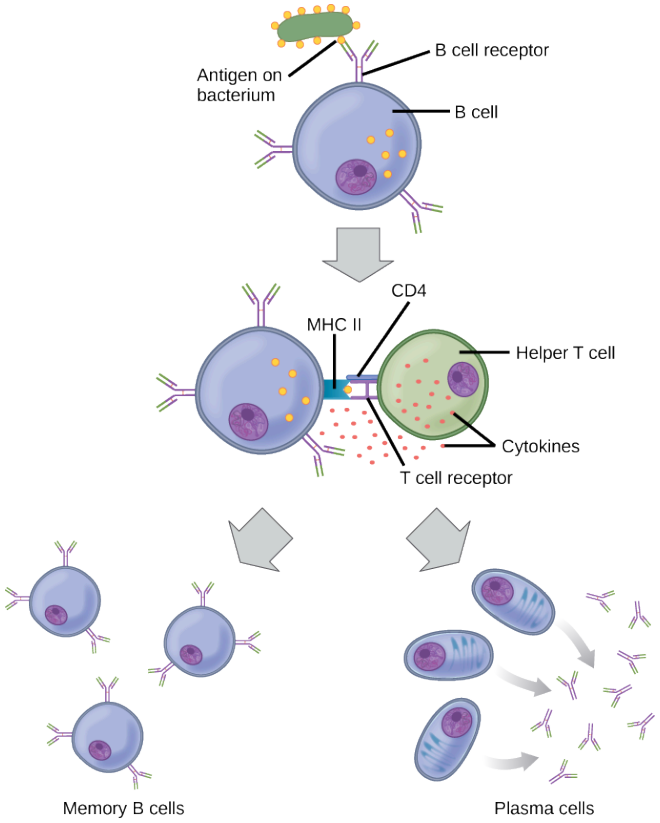


Figure 27.26 After initially binding an antigen to the B cell receptor, a B cell internalizes the antigen and presents it on MHC class II. A helper T cell recognizes the MHC class II- antigen complex and activates the B cell. As a result, memory B cells and plasma cells are made.

If the pathogen is never encountered again during the individual's lifetime, B and T memory cells will circulate for a few years or even several decades and will gradually die off, having never functioned as effector cells. However, if the host is re-exposed to the same pathogen type, circulating memory cells will immediately differentiate into plasma cells and T_C cells without input from APCs or T_H cells. This is known as the secondary immune response. One reason why the adaptive immune response is delayed is because it takes time for naïve B and T cells with the appropriate antigen specificities to be identified, activated, and proliferate. On reinfection, this step is skipped, and the result is a more rapid production of immune defenses. Memory B cells that differentiate into plasma cells produce antibody at a level that is tens to hundreds-fold greater than during the primary response (**Figure 27.27**). This rapid and dramatic antibody response may stop the infection before it can even become established, and the individual may not realize they had been exposed.

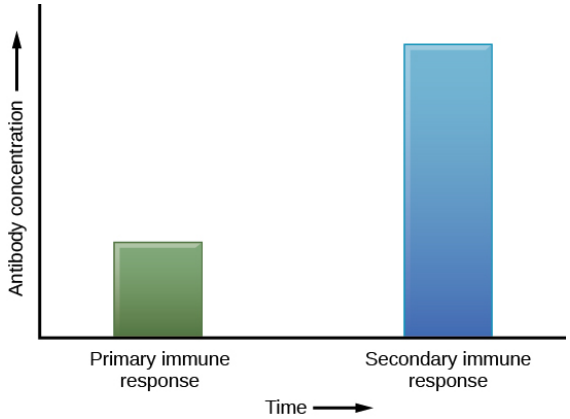


Figure 27.27 In the primary response to infection, antibodies are secreted first from plasma cells. Upon re-exposure to the same pathogen, memory cells differentiate into antibody-secreting plasma cells that output a greater amount of antibody for a longer period of time.

Vaccination is based on the knowledge that exposure to noninfectious antigens, derived from known pathogens, generates a mild primary immune response. The immune response to vaccination may not be perceived by the host as illness but still confers immune memory. When exposed to the corresponding pathogen to which an individual was vaccinated, the reaction is similar to a secondary exposure. Because each reinfection generates more memory cells and increased resistance to the pathogen, some vaccine courses involve one or more booster vaccinations to mimic repeat exposures.

The Lymphatic System

Lymph is the watery fluid that bathes tissues and organs and contains protective white blood cells but does not contain erythrocytes. Lymph moves about the body through the lymphatic system, which is made up of vessels, lymph ducts, lymph glands, and organs, such as tonsils, adenoids, thymus, and spleen.

Although the immune system is characterized by circulating cells throughout the body, the regulation, maturation, and intercommunication of immune factors occur at specific sites. The blood circulates immune cells, proteins, and other factors through the body. Approximately 0.1 percent of all cells in the blood are leukocytes, which include monocytes (the precursor of macrophages) and lymphocytes. Most cells in the blood are red blood cells. Cells of the immune system can travel between the distinct lymphatic and blood circulatory systems, which are separated by interstitial space, by a process called extravasation (passing through to surrounding tissue).

Recall that cells of the immune system originate from stem cells in the bone marrow. B cell maturation occurs in the bone marrow, whereas progenitor cells migrate from the bone marrow and develop and mature into naïve T cells in the organ called the thymus.

On maturation, T and B lymphocytes circulate to various destinations. Lymph nodes scattered throughout the body house large populations of T and B cells, dendritic cells, and macrophages (**Figure 27.28**). Lymph gathers antigens as it drains from tissues. These antigens then are filtered through lymph nodes before the lymph is returned to circulation. APCs in the lymph nodes capture and process antigens and inform nearby lymphocytes about potential pathogens.

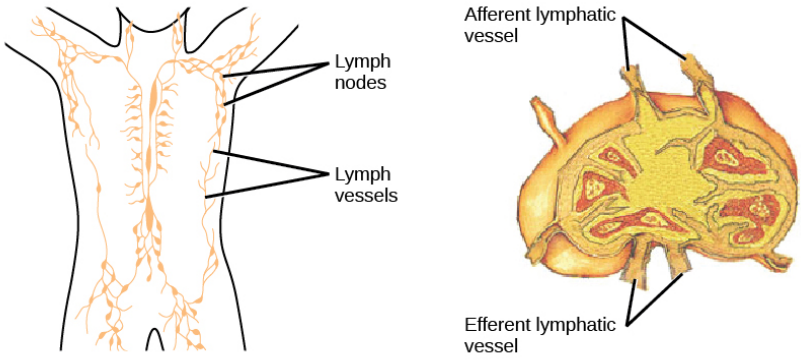


Figure 27.28 (a) Lymphatic vessels carry a clear fluid called lymph throughout the body. The liquid passes through (b) lymph nodes that filter the lymph that enters the node through afferent vessels and leaves through efferent vessels; lymph nodes are filled with lymphocytes that purge infecting cells. (credit a: modification of work by NIH; credit b: modification of work by NCI, NIH)

The spleen houses B and T cells, macrophages, dendritic cells, and NK cells (**Figure 27.29**). The spleen is the site where APCs that have trapped foreign particles in the blood can communicate with lymphocytes. Antibodies are synthesized and secreted by activated plasma cells in the spleen, and the spleen filters foreign substances and antibody-complexed pathogens from the blood. Functionally, the spleen is to the blood as lymph nodes are to the lymph.

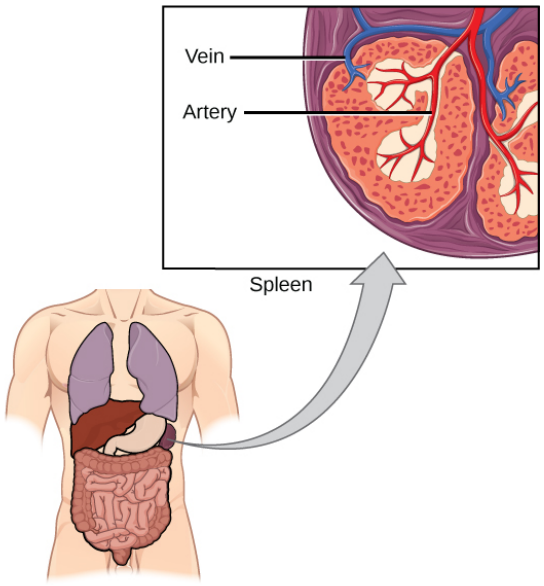


Figure 27.29 The spleen functions to immunologically filter the blood and allow for communication between cells corresponding to the innate and adaptive immune responses. (credit: modification of work by NCI, NIH)

Besides its close ties to and functions with the immune system, the lymphatic system also serves a couple of other important functions. First, the vessels of lymphatic system help absorb excess lymph from body tissues and returns that lymph to the circulatory system via the lymphatic ducts. This helps the body maintain fluid balance. The second function deals with the absorption of lipids during digestion. The lymph vessels located around the small intestine absorb the lacteals from the lumen of the intestine and transport the lacteals to the circulatory system via the lymphatic ducts. In this case the lymphatic system acts as a shuttle for digested fats from the digestive system to the circulatory system.

Immune Tolerance

The immune system has to be regulated to prevent wasteful, unnecessary responses to harmless substances, and more importantly, so that it does not attack "self." The acquired ability to prevent an unnecessary or harmful immune response to a detected foreign substance known not to cause disease, or self-antigens, is described as immune tolerance. The primary mechanism for developing immune tolerance to self-antigens occurs during the selection for weakly self-binding cells during T and B lymphocyte maturation. There are populations of T cells that suppress the immune response to self-antigens and that suppress the immune response after the infection has cleared to minimize host cell damage induced by inflammation and cell lysis. Immune tolerance is especially well developed in the mucosa of the upper digestive system because of the tremendous number of foreign substances (such as food proteins) that APCs of the oral cavity, pharynx, and gastrointestinal mucosa encounter. Immune tolerance is brought about by specialized APCs in the liver, lymph nodes, small intestine, and lung that present harmless antigens to a diverse population of regulatory T (T_{reg}) cells, specialized lymphocytes that suppress local inflammation and inhibit the secretion of stimulatory immune factors. The combined result of T_{reg} cells is to prevent immunologic activation and inflammation in undesired tissue compartments and to allow the immune system to focus on pathogens instead.

27.5 | Systems of Gas Exchange

Introduction

> "The inspired and expired air may be sometimes very useful, by condensing and cooling the blood that passeth through the lungs; I hold that the depuration of the blood in that passage, is not only one of the ordinary, but one of the principal uses of respiration."

Robert Boyle, in *New Experiments ... Touching the Spring of Air*, 1660

The primary function of the respiratory system is to deliver oxygen to the cells of the body's tissues and remove a waste product, carbon dioxide (the process which Boyle called "depuration"). The main structures of the human respiratory system are the nasal cavity, the trachea, and lungs, and these structures are what brings oxygen into the human body and removes carbon dioxide from the human body. As you learned previously, the circulatory system is responsible for moving oxygen from the lungs to the tissues and for moving carbon dioxide from the tissues and taking it to the lungs. At the cellular level the oxygen is needed to make ATP from the energy stored in glucose and other organic molecules; carbon dioxide is a waste product of harvesting that energy. In other words, the respiratory system gets the oxygen inside the body, the circulatory system moves the oxygen around the body getting it to the cells, the cells use the oxygen to produce energy and in the process produce carbon dioxide as a waste, the circulatory system removes the carbon dioxide from the cell and delivers it to the lungs, and the respiratory system removes the carbon dioxide from the body. In vertebrates, the respiratory and circulatory work very closely together in order to allow for gas exchange between the inside and outside of the organism.

All aerobic organisms require oxygen to carry out their metabolic functions. Over evolutionary time, 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 plasma membrane is sufficient for supplying oxygen to the cell (**Figure 27.30**). Diffusion is a slow, passive transport process. 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 systems, to transport oxygen throughout their entire body.

Figure 27.30 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 plasma 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 flat worms, which 'breathe' through diffusion across the outer surface (**Figure 27.31**). The flat shape of these organisms increases the surface area for diffusion, ensuring that each cell within the body is close to the outer 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 27.31 This flatworm's process of respiration works by diffusion across the outer membrane. (credit: Stephen Childs)

Skin and Gills

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 plasma 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 less than that. Fish and many other aquatic organisms have evolved gills to take up the dissolved oxygen from water (**Figure 27.32**). 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 27.32 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 **Figure 27.33**. Similarly, carbon dioxide molecules in the blood diffuse from the blood (high concentration) to water (low concentration).

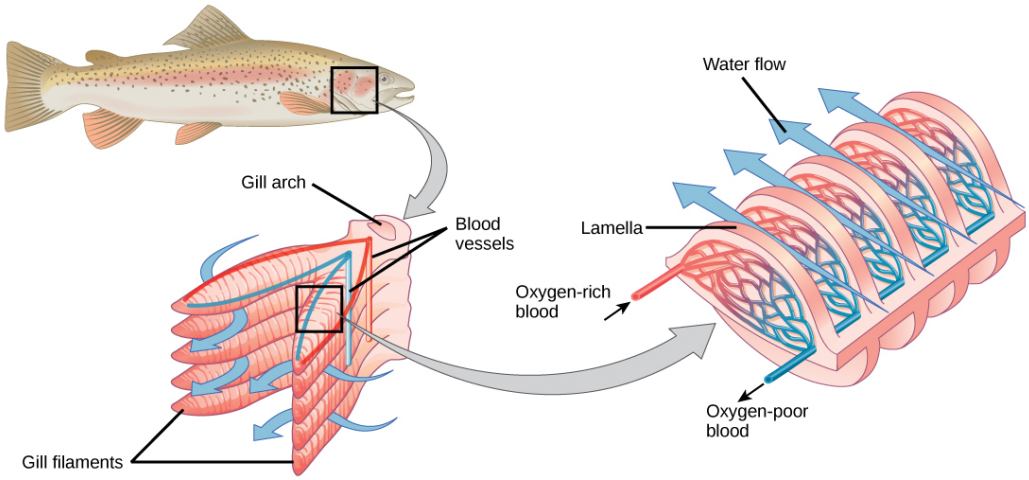


Figure 27.33 As water flows over the gills, oxygen is transferred to blood via the veins. (credit "fish": modification of work by Duane Raver, NOAA)

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.

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 27.34**) 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.

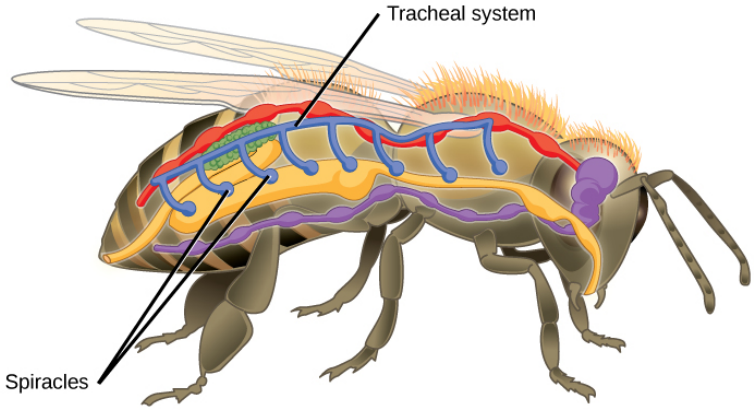


Figure 27.34 Insects perform respiration via a tracheal system.

Mammalian Systems

In mammals, pulmonary ventilation occurs via inhalation (breathing). During inhalation, air enters the human body through the **nasal cavity** located just inside the nose (**Figure 27.35**). 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.

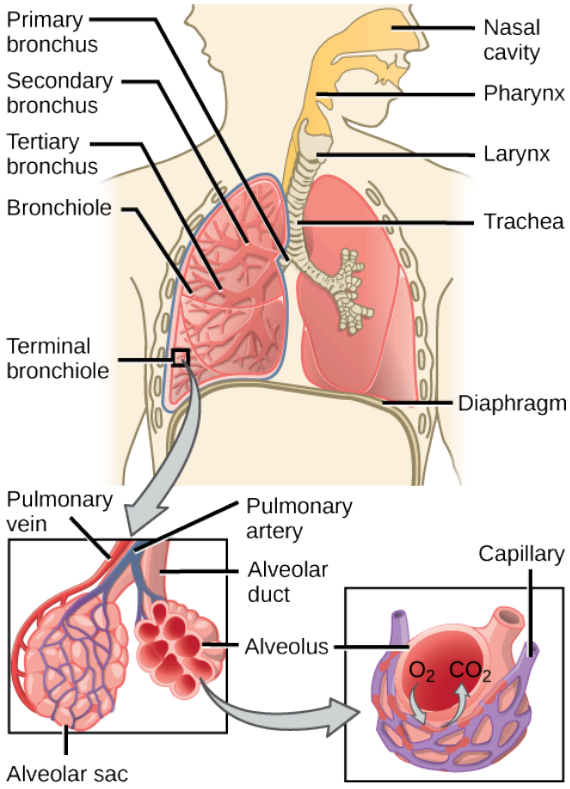


Figure 27.35 Air enters the respiratory system through the nasal cavity and pharynx, and then passes through the trachea and into the bronchi, which bring air into the lungs. (credit: modification of work by NCI)

From the nasal cavity, air passes through the **pharynx** (throat) and the **larynx** (voice box), as it makes its way to the **trachea** (**Figure 27.35**). 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 midthorax (**Figure 27.35**). 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 expel mucus when we cough. Smooth muscle can contract or relax, depending on stimuli from the external environment or the body's nervous system.

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 27.36**). The muscular **diaphragm**, which facilitates breathing, is inferior (below) to the lungs and marks the end of the thoracic cavity.

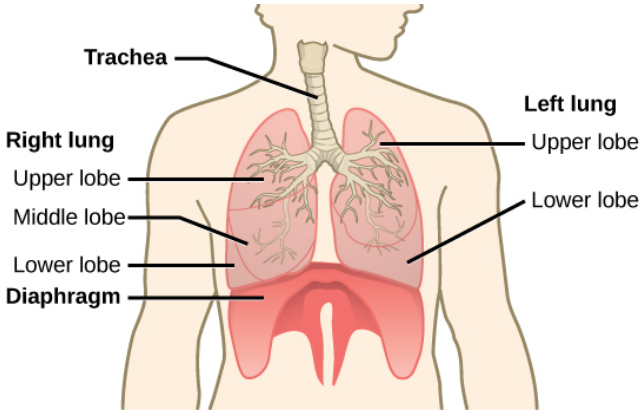


Figure 27.36 The trachea bifurcates into the right and left bronchi in the lungs. The right lung is made of three lobes and is larger. To accommodate the heart, the left lung is smaller and has only two lobes.

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.

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 27.37**). 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.

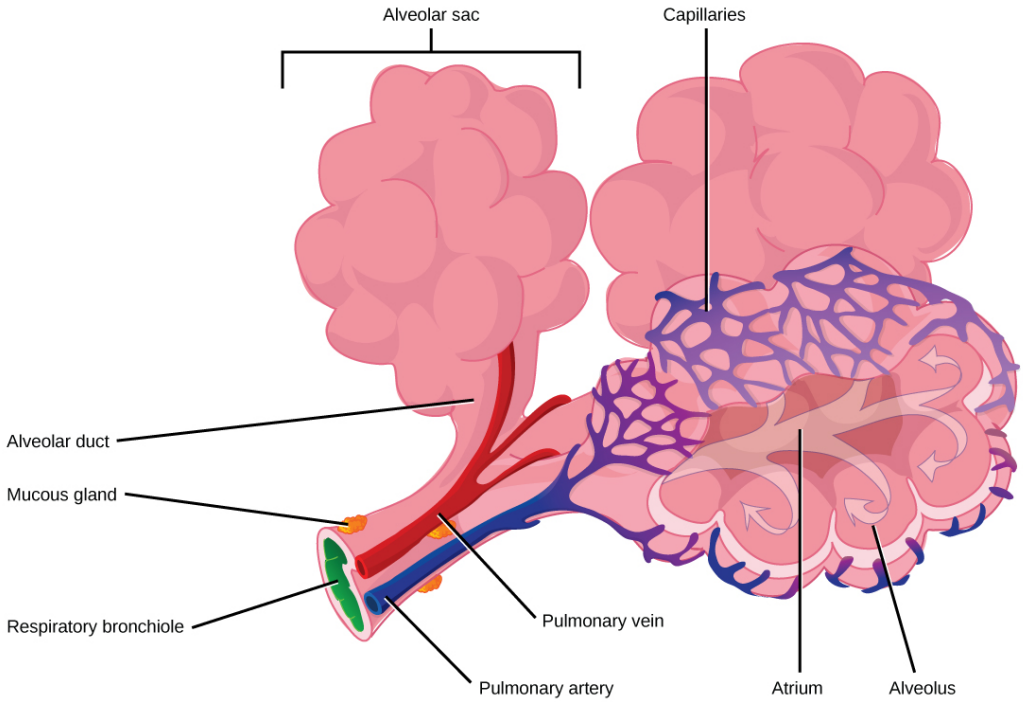


Figure 27.37 Terminal bronchioles are connected by respiratory bronchioles to alveolar ducts and alveolar sacs. Each alveolar sac contains 20 to 30 spherical alveoli and has the appearance of a bunch of grapes. Air flows into the atrium of the alveolar sac, then circulates into alveoli where gas exchange occurs with the capillaries. Mucous glands secrete mucous into the airways, keeping them moist and flexible. (credit: modification of work by Mariana Ruiz Villareal)

Transport of Gases in Blood

Once the oxygen diffuses across the alveoli, it enters the bloodstream and is transported to the tissues where it is unloaded,

and carbon dioxide diffuses out of the blood and into the alveoli to be expelled from the body. Although gas exchange is a continuous process, the oxygen and carbon dioxide are transported by different mechanisms.

Transport of Oxygen in the Blood

Although oxygen dissolves in blood, only a small amount of oxygen is transported this way. Only 1.5 percent of oxygen in the blood is dissolved directly into the blood itself. Most oxygen—98.5 percent—is bound to a protein called hemoglobin and carried to the tissues.

Hemoglobin

Hemoglobin, or Hb, is a protein molecule found in red blood cells (erythrocytes) made of four subunits: two alpha subunits and two beta subunits (**Figure 27.38**). Each subunit surrounds a central **heme group** that contains iron and binds one oxygen molecule, allowing each hemoglobin molecule to bind four oxygen molecules. Molecules with more oxygen bound to the heme groups are brighter red. As a result, oxygenated arterial blood where the Hb is carrying four oxygen molecules is bright red, while venous blood that is deoxygenated is darker red.

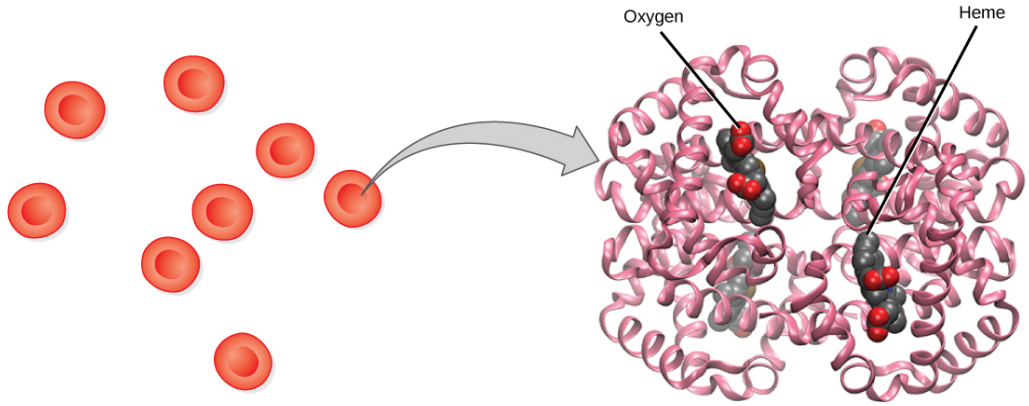


Figure 27.38 The protein inside (a) red blood cells that carries oxygen to cells and carbon dioxide to the lungs is (b) hemoglobin. Hemoglobin is made up of four symmetrical subunits and four heme groups. Iron associated with the heme binds oxygen. It is the iron in hemoglobin that gives blood its red color.

Transport of Carbon Dioxide in the Blood

Carbon dioxide molecules are transported in the blood from body tissues to the lungs by one of three methods: dissolution directly into the blood, binding to hemoglobin, or carried as a bicarbonate ion. Several properties of carbon dioxide in the blood affect its transport. First, carbon dioxide is more soluble in blood than oxygen. About 5 to 7 percent of all carbon dioxide is dissolved in the plasma. Second, carbon dioxide can bind to plasma proteins or can enter red blood cells and bind to hemoglobin. This form transports about 10 percent of the carbon dioxide. When carbon dioxide binds to hemoglobin, a molecule called carbaminohemoglobin is formed. Binding of carbon dioxide to hemoglobin is reversible. Therefore, when it reaches the lungs, the carbon dioxide can freely dissociate from the hemoglobin and be expelled from the body.

Third, the majority of carbon dioxide molecules (85 percent) are carried as part of the bicarbonate buffer system. In this system, carbon dioxide diffuses into the red blood cells. Carbonic anhydrase (CA) within the red blood cells quickly converts the carbon dioxide into carbonic acid (H_2CO_3). Carbonic acid is an unstable intermediate molecule that immediately dissociates into bicarbonate ions (HCO_3^-) and hydrogen (H^+) ions. Since carbon dioxide is quickly converted into bicarbonate ions, this reaction allows for the continued uptake of carbon dioxide into the blood down its concentration gradient. It also results in the production of H^+ ions. If too much H^+ is produced, it can alter blood pH. However, hemoglobin binds to the free H^+ ions and thus limits shifts in pH. The newly synthesized bicarbonate ion is transported out of the red blood cell into the plasma of the blood in exchange for a chloride ion (Cl^-); this is called the chloride shift. When the blood reaches the lungs, the bicarbonate ion is transported back into the red blood cell in exchange for the chloride ion. The H^+ ion dissociates from the hemoglobin and binds to the bicarbonate ion. This produces the carbonic acid intermediate, which is converted back into carbon dioxide through the enzymatic action of CA. The carbon dioxide produced is expelled through the lungs during exhalation.

$$CO_2 + H_2O \leftrightarrow \underset{\text{(carbonic acid)}}{H_2CO_3} \leftrightarrow \underset{\text{(bicarbonate)}}{HCO_3 + H^+}$$

The benefit of the bicarbonate buffer system is that carbon dioxide is “soaked up” into the blood with little change to the pH of the system. This is important because it takes only a small change in the overall pH of the body for severe injury or death to result. The presence of this bicarbonate buffer system also allows for people to travel and live at high altitudes: When the partial pressure of oxygen and carbon dioxide change at high altitudes, the bicarbonate buffer system adjusts to regulate carbon dioxide while maintaining the correct pH in the body.

28 | ENDOCRINE, REPRODUCTIVE AND URINARY SYSTEMS

28.1 | Endocrine System

> "The specific character of the greater part of the toxins which are known to us (I need only instance such toxins as those of tetanus and diphtheria) would suggest that the substances produced for effecting the corrleation of organs within the body, through the intermediation of the blood stream, might also belong to this class, since here also **specificity of action** must be a distinguishing characteristic. These chemical messengers, however, or "hormones"(from the Greek ὁρμῶν, to excite or arouse), as we might call them, have to be carried from the organ where they are produced to the organ which they affect by means of the blood stream, and the continually recurring physiological needs of the organism must determine their repeated production and circulation throughout the body. "

Ernest Henry Starling, "The Chemical Correlation of the Functions of the Body", *The Lancet*, 1905, II, 340

Hormones, as Starling noted, are produced by one organ and affect the activities of other organs. Unlike **neurotransmitters**, which you will learn about later in this module, hormones move via the bloodstream from the site of production to the site of action. But like neurotransmitters, hormones are key players in maintaining homeostasis. Before we discuss that, however, we need to review homeostasis and introduce the major classes of animal hormones.

Types of Hormones

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 chemicals that are released by cells into body fluids (usually blood) and which act on target cells at some distance from the cells that release the hormone. At the **target cells**, which are cells that have a receptor for the chemical, 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.

Although there are many different hormones in the human body, they can be divided into two general classes based on their chemical structure and water solubility: steroid hormones (most are derivatives of cholesterol), which are not soluble in water, and peptide (peptides and proteins) hormones, which are readily soluble in water. One of the key distinguishing features of lipid-derived hormones is that they can diffuse across plasma membranes whereas the peptide hormones cannot.

Lipid-Derived Hormones (or Lipid-soluble Hormones)

Most **lipid hormones** are derived from cholesterol and thus are structurally similar to it, as illustrated in **Figure 28.1**. The primary class of lipid hormones in humans is the steroid hormones. Examples of steroid hormones include estradiol, which is an **estrogen**, or female sex hormone, and testosterone, which is an androgen, or male sex hormone. These two hormones are released by the female and male reproductive organs, respectively. Other steroid hormones include aldosterone and cortisol, which are released by the adrenal glands along with some other types of androgens. Steroid hormones are insoluble in water, and need to be bound to transport proteins in order to be transported in the blood. As a result, they remain in the body longer than peptide hormones. For example, cortisol has a half-life of 60 to 90 minutes in humans, while epinephrine, an amino acid derived-hormone, has a half-life of approximately one minute.

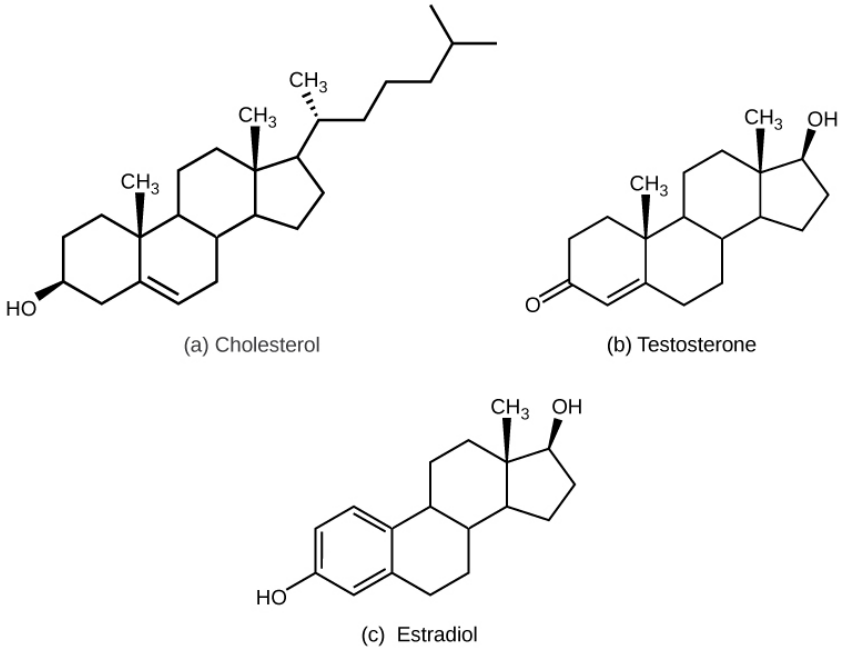


Figure 28.1 The structures shown here represent (a) cholesterol, plus the steroid hormones (b) testosterone and (c) estradiol.

Peptide (water-soluble) Hormones

The **peptide hormones** include polypeptides as well as several relatively small molecules that are derived from the amino acids tyrosine and tryptophan, shown in **Figure 28.2**. Examples of amino acid-derived hormones include epinephrine and norepinephrine, which are synthesized in the medulla of the adrenal glands, and thyroxine, which is produced by the thyroid gland. The pineal gland in the brain makes and secretes melatonin which regulates sleep cycles.

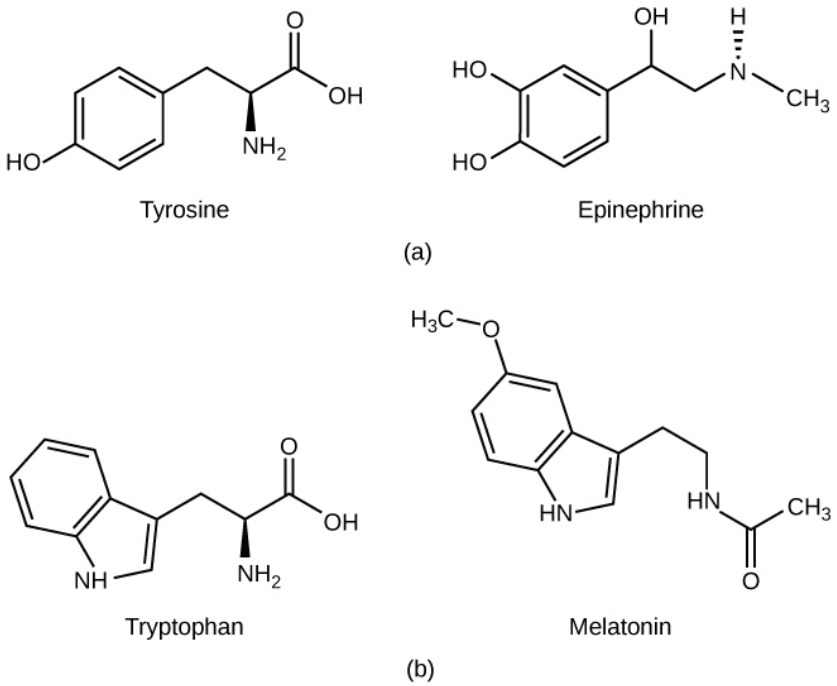


Figure 28.2 (a) The hormone epinephrine, which triggers the fight-or-flight response, is derived from the amino acid tyrosine. (b) The hormone melatonin, which regulates circadian rhythms, is derived from the amino acid tryptophan.

Other **peptide hormones** are polypeptides (chains of amino acids linked by peptide bonds). These hormones include molecules that are quite short polypeptide chains, such as antidiuretic hormone (9 amino acids) and oxytocin (also 9 amino acids), both of which are produced in the brain and released into the blood in the posterior pituitary gland. This class also includes small proteins, like the growth hormones (approx 190 amino acids(produced by the pituitary, and large glycoproteins such as follicle-stimulating hormone (a complex of 2 different polypeptides, each about 100 amino acids in length), produced by the pituitary. **Figure 28.3** illustrates these peptide hormones.

Secreted peptides like insulin are stored within vesicles in the cells that synthesize them. They are then released in response to stimuli such as high blood glucose levels in the case of insulin. Amino acid-derived and polypeptide hormones are water-soluble. Therefore these hormones cannot cross the plasma membranes of cells; their receptors are found on the surface of the target cells.

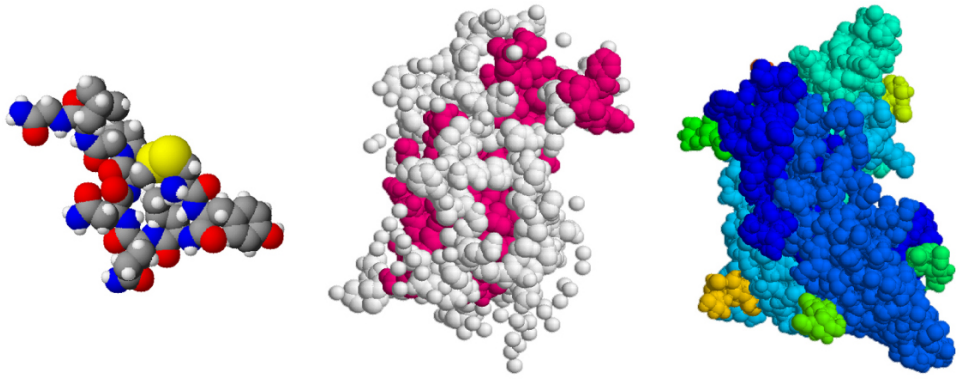


Figure 28.3 The structures of peptide hormones (a) oxytocin, (b) growth hormone, and (c) follicle-stimulating hormone are shown. These peptide hormones are much larger than those derived from cholesterol or amino acids.

How Hormones Work

Hormones mediate changes in target cells after 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; testosterone receptors are found in relatively few cell types. 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.

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-soluble hormones such as steroid hormones diffuse across the membranes of the cells where they are produced. 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 bind 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. The cellular responses are varied, ranging from changes in the structure of the cell to the production of enzymes that catalyze new chemical reactions. In this way, the steroid hormone regulates specific cell processes as illustrated in **Figure 28.4**.

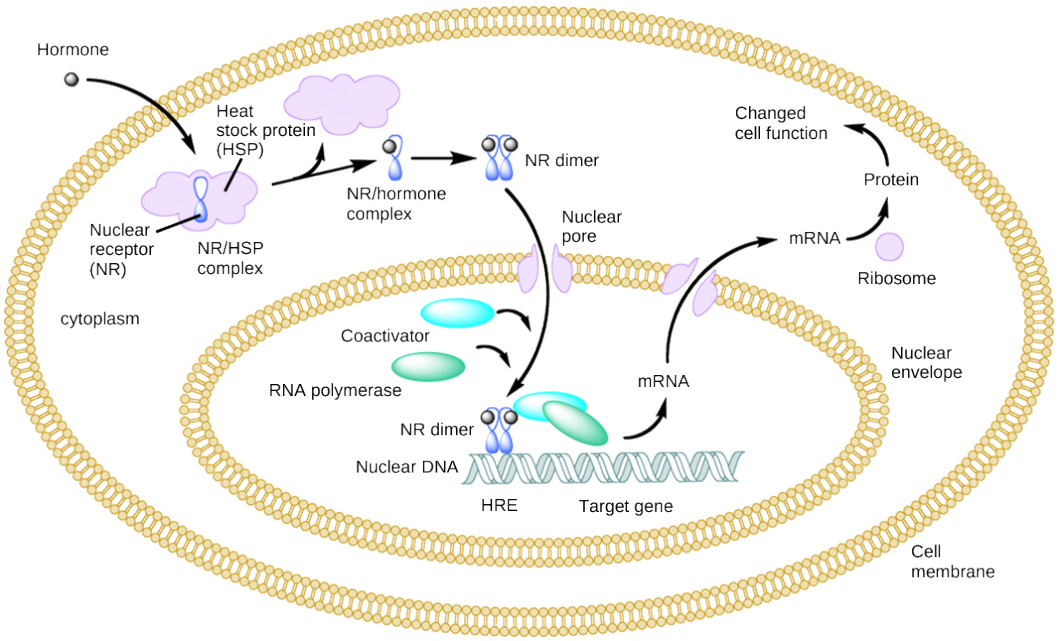


Figure 28.4 An intracellular nuclear receptor (NR) is located in the cytoplasm bound to a heat shock protein (HSP). Upon hormone binding, the receptor dissociates from the heat shock protein and translocates to the nucleus. In the nucleus, the hormone-receptor complex binds to a DNA sequence called a hormone response element (HRE), which triggers gene transcription and translation. The corresponding protein product can then mediate changes in cell function.

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

Peptide hormones are not 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. 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 plasma 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 28.5**.

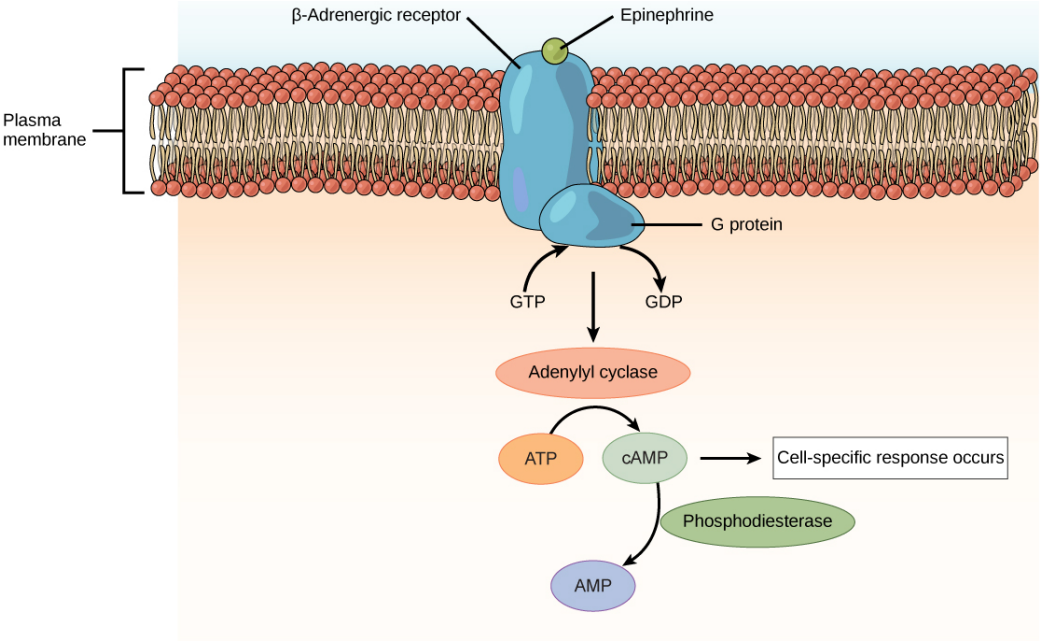


Figure 28.5 The amino acid-derived hormones epinephrine and norepinephrine bind to beta-adrenergic receptors on the plasma membrane of cells. Hormone binding to receptor activates a G-protein, which in turn activates adenylyl cyclase, converting ATP to cAMP. cAMP is a second messenger that mediates a cell-specific response. An enzyme called phosphodiesterase breaks down cAMP, terminating the signal.

The specific response of a cell to a lipid insoluble hormone depends on the type of receptors that are present on the plasma 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.

Hormonal Regulation of Body Systems

Hormones have a wide range of effects and modulate many different body processes. Two regulatory processes that will be examined here as examples are regulation of the functions of the reproductive system, and regulation of carbohydrate metabolism.

Hormonal Regulation of the Reproductive System

Regulation of the reproductive system is a process that requires the action of hormones from the pituitary gland, the adrenal cortex, and the gonads. During puberty in both males and females, the hypothalamus produces gonadotropin-releasing hormone (GnRH), which stimulates the production and release of **follicle-stimulating hormone (FSH)** and luteinizing hormone (LH) from the anterior pituitary gland. These hormones regulate the gonads (testes in males and ovaries in females) and therefore are called **gonadotropins**. In both males and females, FSH stimulates gamete production and LH stimulates production of hormones by the gonads. An increase in gonad hormone levels inhibits GnRH production through a negative feedback loop.

Regulation of the Male Reproductive System

In males, FSH stimulates the maturation of sperm cells. FSH production is inhibited by the hormone inhibin, which is released by the testes. LH stimulates production of the sex hormones (**androgens**) by the interstitial cells of the testes and therefore is also called interstitial cell-stimulating hormone.

The most widely known androgen in males is testosterone. Testosterone promotes the production of sperm and masculine characteristics.

everyday CONNECTION

The Dangers of Synthetic Hormones

Figure 28.6 Professional baseball player Jason Giambi publically admitted to, and apologized for, his use of anabolic steroids supplied by a trainer. (credit: Bryce Edwards)

Some athletes attempt to boost their performance by using artificial hormones that enhance muscle performance. Anabolic steroids, a form of the male sex hormone testosterone, are one of the most widely known performance-enhancing drugs. Steroids are used to help build muscle mass. Other hormones that are used to enhance athletic performance include erythropoietin, which triggers the production of red blood cells, and human growth hormone, which can help in building muscle mass. Most performance enhancing drugs are illegal for non-medical purposes. They are also banned by national and international governing bodies including the International Olympic Committee, the U.S. Olympic Committee, the National Collegiate Athletic Association, the Major League Baseball, and the National Football League.

The side effects of synthetic hormones are often significant and non-reversible, and in some cases, fatal. Androgens produce several complications such as liver dysfunctions and liver tumors, prostate gland enlargement, difficulty urinating, premature closure of epiphyseal cartilages, testicular atrophy, infertility, and immune system depression. The physiological strain caused by these substances is often greater than what the body can handle, leading to unpredictable and dangerous effects and linking their use to heart attacks, strokes, and impaired cardiac function.

Regulation of the Female Reproductive System

In females, FSH stimulates development of egg cells, called ova, which develop in structures called follicles. Follicle cells produce the hormone inhibin, which inhibits FSH production. LH also plays a role in the development of ova, induction of ovulation, and stimulation of estradiol and progesterone production by the ovaries, as illustrated in **Figure 28.7**. Estradiol and progesterone are steroid hormones that prepare the body for pregnancy. Estradiol produces secondary sex characteristics in females, while both estradiol and progesterone regulate the menstrual cycle.

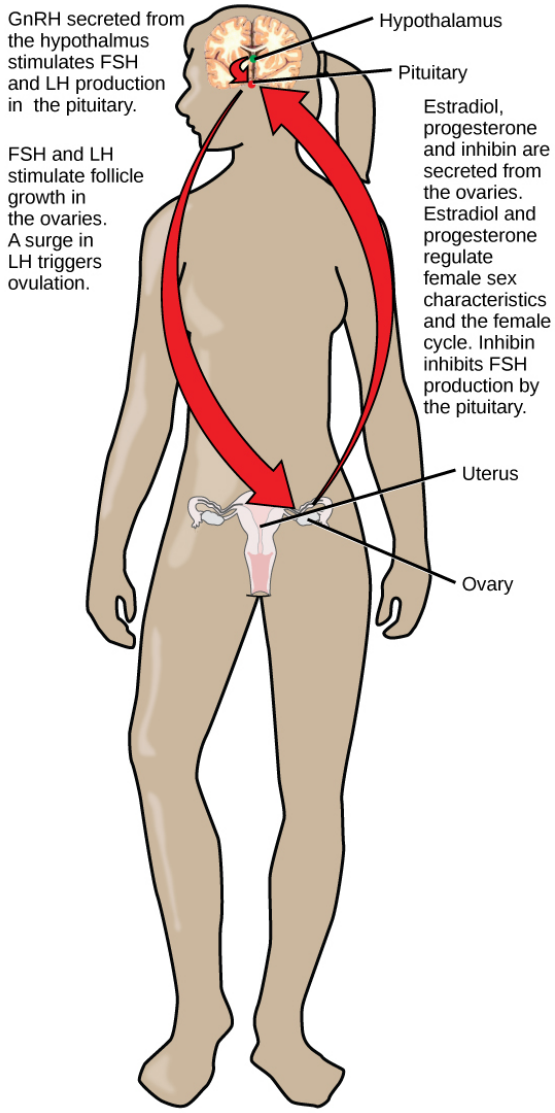


Figure 28.7 Hormonal regulation of the female reproductive system involves hormones from the hypothalamus, pituitary, and ovaries.

In addition to producing FSH and LH, the anterior portion of the pituitary gland also produces the hormone prolactin (PRL) in females. Prolactin stimulates the production of milk by the mammary glands following childbirth. Prolactin levels are regulated by the hypothalamic hormones prolactin-releasing hormone (PRH) and prolactin-inhibiting hormone (PIH), which is now known to be dopamine. PRH stimulates the release of prolactin and PIH inhibits it; this is a classic negative feedback loop.

The posterior pituitary releases the hormone oxytocin, which stimulates uterine contractions during childbirth. The uterine smooth muscles are not very sensitive to oxytocin until late in pregnancy when the number of oxytocin receptors in the uterus peaks. Stretching of tissues in the uterus and cervix stimulates oxytocin release during childbirth. Contractions increase in intensity as blood levels of oxytocin rise via a positive feedback mechanism until the birth is complete. Oxytocin also stimulates the contraction of myoepithelial cells around the milk-producing mammary glands. As these cells contract, milk is forced from the secretory alveoli into milk ducts and is ejected from the breasts in milk ejection (“let-down”) reflex. Oxytocin release is stimulated by the suckling of an infant, which triggers the synthesis of oxytocin in the hypothalamus and its release into circulation at the posterior pituitary.

Hormonal Regulation of Carbohydrate Metabolism

Blood glucose levels vary widely over the course of a day as periods of food consumption alternate with periods of fasting. Insulin and glucagon are the two hormones primarily responsible for maintaining homeostasis of blood glucose levels. Additional regulation is mediated by the thyroid hormones.

Regulation of Blood Glucose Levels by Insulin and Glucagon

Cells of the body require nutrients in order to function, and these nutrients are obtained through feeding. In order to manage nutrient intake, storing excess intake and utilizing reserves when necessary, the body uses hormones to moderate energy stores. **Insulin** is produced by the beta cells of the pancreas, which are stimulated to release insulin as blood glucose levels rise (for example, after a meal is consumed). Insulin lowers blood glucose levels by enhancing the rate of glucose uptake and utilization by target cells, which use glucose for ATP production. It also stimulates the liver to convert glucose to glycogen, which is then stored by cells for later use. Insulin also increases glucose transport into certain cells, such as muscle cells, fat cells, and liver cells. This results from an insulin-mediated increase in the number of glucose transporter proteins in plasma membranes, which remove glucose from circulation by facilitated diffusion. As insulin binds to its target cell via insulin receptors and signal transduction, it triggers the cell to incorporate glucose transport proteins into its membrane. Insulin also stimulates the conversion of glucose to fat in adipocytes and the synthesis of proteins. These actions mediated by insulin cause blood glucose concentrations to fall, called a hypoglycemic "low sugar" effect, which inhibits further insulin release from beta cells through a negative feedback loop.

When blood glucose levels decline below normal levels, for example between meals or when glucose is utilized rapidly during exercise, the hormone **glucagon** is released from the alpha cells of the pancreas. Glucagon raises blood glucose levels, eliciting what is called a hyperglycemic effect, by stimulating the breakdown of glycogen to glucose in skeletal muscle cells and liver cells in a process called glycogenolysis. Glucose can then be utilized as energy by muscle cells and released into circulation by the liver cells. Glucagon also stimulates absorption of amino acids from the blood by the liver, which then converts them to glucose. This process of glucose synthesis is called gluconeogenesis. Glucagon also stimulates adipose cells to release fatty acids into the blood. These actions mediated by glucagon result in an increase in blood glucose levels to normal homeostatic levels. Rising blood glucose levels inhibit further glucagon release by the pancreas via a negative feedback mechanism. In this way, insulin and glucagon work together to maintain homeostatic glucose levels, as shown in **m47442 (http://legacy.cnx.org/content/m47442/1.6/#fig-ch37_03_05)** .

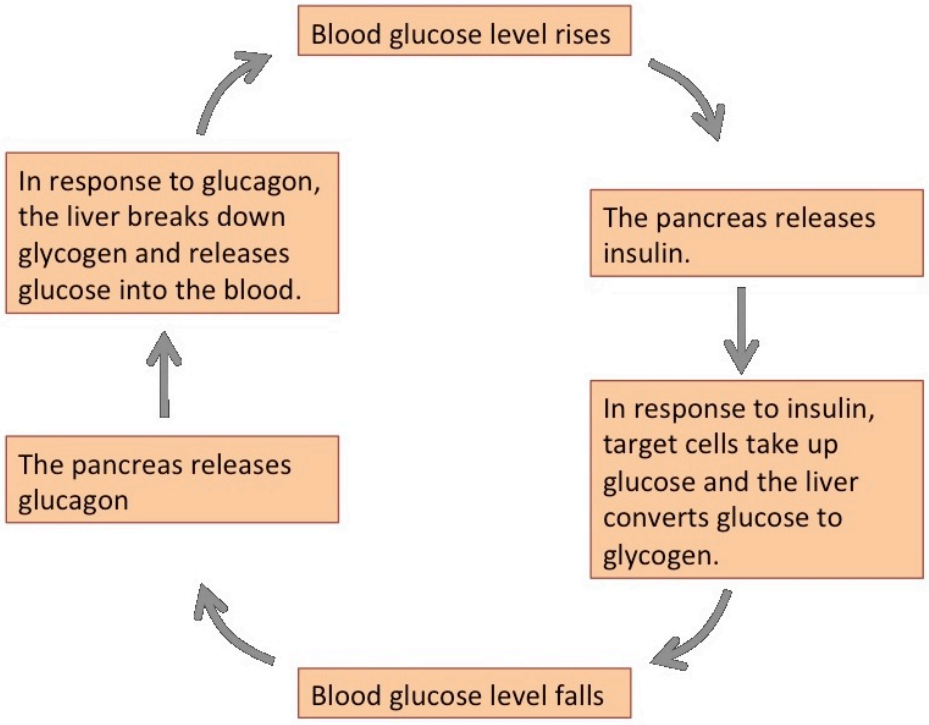


Figure 28.8 Insulin and glucagon regulate blood glucose levels via negative feedback mechanisms.

Impaired insulin function can lead to a condition called **diabetes mellitus**, the main symptoms of which are illustrated in **Figure 28.9**. This can be caused by low levels of insulin production by the beta cells of the pancreas, or by reduced sensitivity of tissue cells to insulin. This prevents glucose from being absorbed by cells, causing high levels of blood glucose, or **hyperglycemia** (high sugar). High blood glucose levels make it difficult for the kidneys to recover all the glucose from nascent urine, resulting in glucose being lost in urine. High glucose levels also result in less water being reabsorbed by the kidneys, causing high amounts of urine to be produced; this may result in dehydration. Over time, high

blood glucose levels can cause nerve damage to the eyes and peripheral body tissues, as well as damage to the kidneys and cardiovascular system. Oversecretion of insulin can cause **hypoglycemia**, low blood glucose levels. This causes insufficient glucose availability to cells, often leading to muscle weakness, and can sometimes cause unconsciousness or death if left untreated.

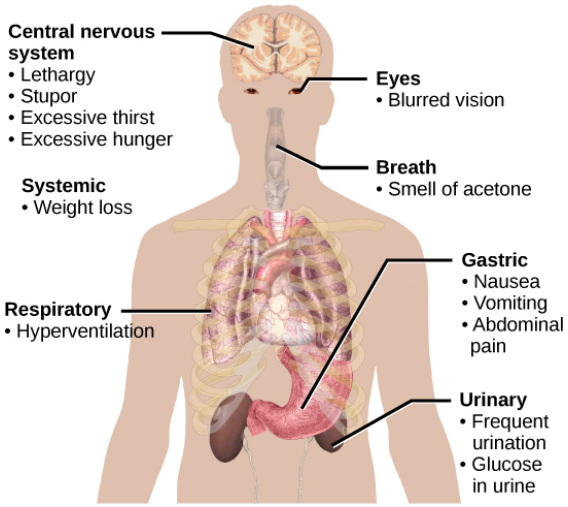


Figure 28.9 The main symptoms of diabetes are shown. (credit: modification of work by Mikael Häggström)

Endocrine Glands

Both the endocrine and nervous systems use chemical signals to communicate and regulate the body's physiology. The endocrine system releases hormones that act on target cells to regulate development, growth, energy metabolism, reproduction, and many behaviors. The nervous system releases neurotransmitters or neurohormones that regulate neurons, muscle cells, and endocrine cells. Because the neurons can regulate the release of hormones, the nervous and endocrine systems work in a coordinated manner to regulate the body's physiology.

Hypothalamic-Pituitary Axis

The **hypothalamus** in vertebrates integrates the endocrine and nervous systems. The hypothalamus is an endocrine organ located in the diencephalon of the brain. It receives input from the body and other brain areas and initiates endocrine responses to environmental changes. The hypothalamus acts as an endocrine organ, synthesizing hormones and transporting them along axons to the posterior pituitary gland. It synthesizes and secretes regulatory hormones that control the endocrine cells in the anterior pituitary gland. The hypothalamus contains autonomic centers that control endocrine cells in the adrenal medulla via neuronal control.

The **pituitary gland**, sometimes called the hypophysis or “master gland” is located at the base of the brain in the sella turcica, a groove of the sphenoid bone of the skull, illustrated in **Figure 28.10**. It is attached to the hypothalamus via a stalk called the **pituitary stalk** (or infundibulum). The anterior portion of the pituitary gland is regulated by releasing or release-inhibiting hormones produced by the hypothalamus, and the posterior pituitary receives signals via neurosecretory cells to release hormones produced by the hypothalamus. The pituitary has two distinct regions—the anterior pituitary and the posterior pituitary—which between them secrete nine different peptide or protein hormones. The posterior lobe of the pituitary gland contains axons of the hypothalamic neurons.

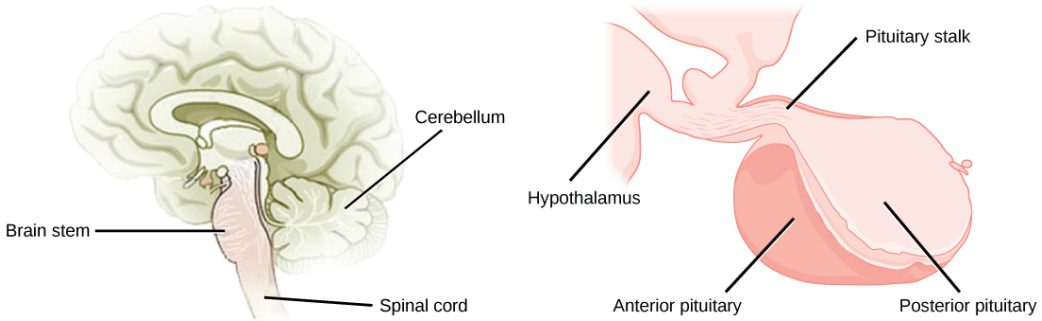


Figure 28.10 The pituitary gland is located at (a) the base of the brain and (b) connected to the hypothalamus by the pituitary stalk. (credit a: modification of work by NCI; credit b: modification of work by Gray's Anatomy)

Thyroid Gland

The **thyroid gland** is located in the neck, just below the larynx and in front of the trachea, as shown in **Figure 28.11**. It is a butterfly-shaped gland with two lobes that are connected by the isthmus. It has a dark red color due to its extensive vascular system. When the thyroid swells due to dysfunction, it can be felt under the skin of the neck.

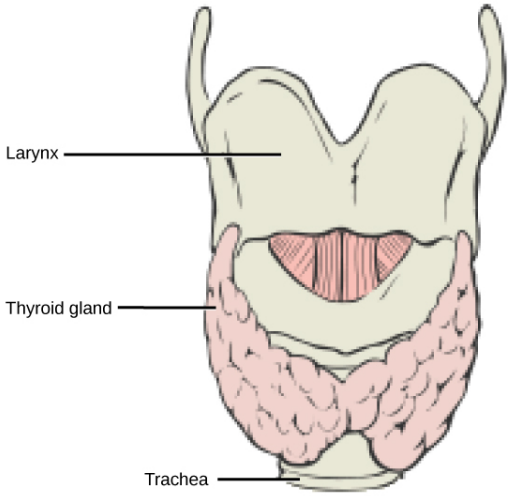


Figure 28.11 This illustration shows the location of the thyroid gland.

Thyroid follicle cells synthesize the hormone thyroxine, which is also known as T_4 because it contains four atoms of iodine, and triiodothyronine, also known as T_3 because it contains three atoms of iodine. Follicle cells are stimulated to release stored T_3 and T_4 by thyroid stimulating hormone (TSH), which is produced by the anterior pituitary. These thyroid hormones increase the rates of mitochondrial ATP production.

A third hormone, calcitonin, is produced by parafollicular cells of the thyroid either releasing hormones or inhibiting hormones. Calcitonin release is not controlled by TSH, but instead is released when calcium ion concentrations in the blood rise. Calcitonin functions to help regulate calcium concentrations in body fluids. It acts in the bones to inhibit osteoclast activity and in the kidneys to stimulate excretion of calcium. The combination of these two events lowers body fluid levels of calcium.

Parathyroid Glands

Most people have four **parathyroid glands**; however, the number can vary from two to six. These glands are located on the posterior surface of the thyroid gland, as shown in **Figure 28.12**. Normally, there is a superior gland and an inferior gland associated with each of the thyroid's two lobes. Each parathyroid gland is covered by connective tissue and contains many secretory cells that are associated with a capillary network.

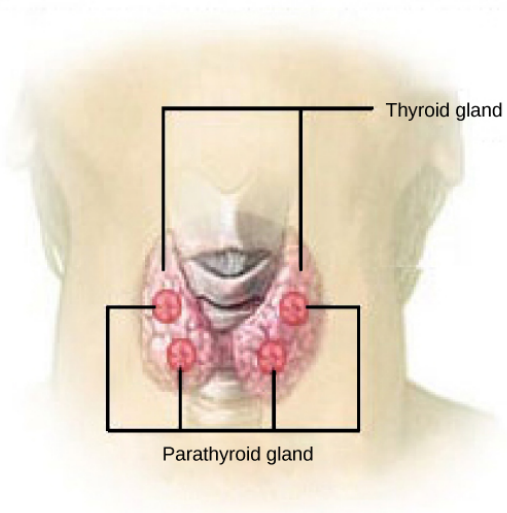


Figure 28.12 The parathyroid glands are located on the posterior of the thyroid gland. (credit: modification of work by NCI)

The parathyroid glands produce parathyroid hormone (PTH). PTH increases blood calcium concentrations when calcium ion levels fall below normal. PTH (1) enhances reabsorption of Ca^{2+} by the kidneys, (2) stimulates osteoclast activity and inhibits osteoblast activity, and (3) it stimulates synthesis and secretion of calcitriol by the kidneys, which enhances Ca^{2+} absorption by the digestive system. PTH and calcitonin work in opposition to one another to maintain homeostatic Ca^{2+} levels in body fluids.

Adrenal Glands

The **adrenal glands** are associated with the kidneys; one gland is located on top of each kidney as illustrated in **Figure 28.13**. The adrenal glands consist of an outer adrenal cortex and an inner adrenal medulla. These regions secrete different hormones.

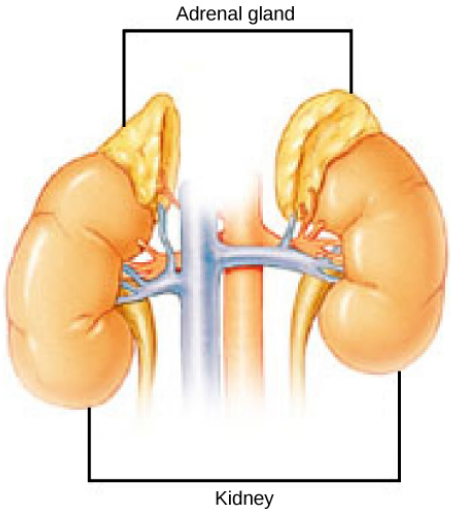


Figure 28.13 The location of the adrenal glands on top of the kidneys is shown. (credit: modification of work by NCI)

Adrenal Cortex

The **adrenal cortex** is made up of layers of epithelial cells and associated capillary networks. This gland produces mineralcorticoids, glucocorticoids, and androgens. The main **mineralocorticoid** (a class of steroid hormones that regulate salt and water balance) is aldosterone, which regulates the concentration of Na^{+} ions in urine, sweat, pancreas, and saliva. Aldosterone release from the adrenal cortex is stimulated by a decrease in blood concentrations of sodium ions, blood

volume, or blood pressure, or by an increase in blood potassium levels.

The three main **glucocorticoids** (steroid hormones that regulate glucose metabolism) are cortisol, corticosterone, and cortisone. The glucocorticoids stimulate the synthesis of glucose. and can also enhance gluconeogenesis (conversion of a non-carbohydrate to glucose) by liver cells. They also promote the release of fatty acids from adipose tissue. These hormones increase blood glucose levels to maintain levels within a normal range between meals. These hormones are secreted in response to ACTH, and levels are regulated by negative feedback.

Androgens are sex hormones that promote masculinity. They are produced in small amounts by the adrenal cortex in both males and females. They do not affect sexual characteristics and may supplement sex hormones released from the gonads.

Adrenal Medulla

The adrenal medulla contains two types of secretory cells: one that produces epinephrine (adrenaline) and another that produces norepinephrine (noradrenaline). Epinephrine is the primary adrenal medulla hormone accounting for 75 to 80 percent of its secretions. Epinephrine and norepinephrine increase heart rate, breathing rate, cardiac muscle contractions, blood pressure, and blood glucose levels. They also accelerate the breakdown of glucose in skeletal muscles and stored fats in adipose tissue.

The release of epinephrine and norepinephrine is stimulated by neural impulses from the sympathetic nervous system. Neural impulses, originating from the hypothalamus in response to stress, release these hormones to prepare the body for the fight-or-flight response.

Pancreas

The **pancreas**, illustrated in Figure 28.14, is an elongated organ that is located between the stomach and the proximal portion of the small intestine. It contains both exocrine cells that excrete digestive enzymes and endocrine cells that release hormones. It is sometimes referred to as a heterocrine gland because it has both endocrine and exocrine functions.

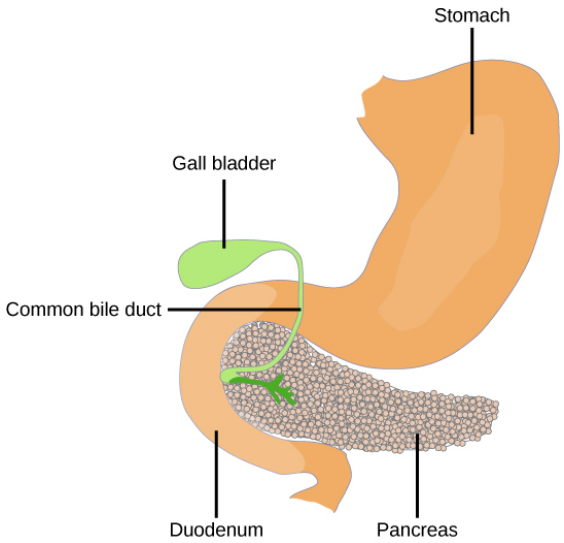


Figure 28.14 The pancreas is found underneath the stomach and points toward the spleen. (credit: modification of work by NCI)

The endocrine cells of the pancreas form clusters called pancreatic islets or the **islets of Langerhans**, as visible in the micrograph shown in Figure 28.15. The pancreatic islets contain two primary cell types: **alpha cells**, which produce the hormone glucagon, and **beta cells**, which produce the hormone insulin. These hormones regulate blood glucose levels. As blood glucose levels decline, alpha cells release glucagon to raise the blood glucose levels by increasing rates of glycogen breakdown and glucose release by the liver. When blood glucose levels rise, such as after a meal, beta cells release insulin to lower blood glucose levels by increasing the rate of glucose uptake in most body cells, and by increasing glycogen synthesis in skeletal muscles and the liver. Together, glucagon and insulin regulate blood glucose levels.

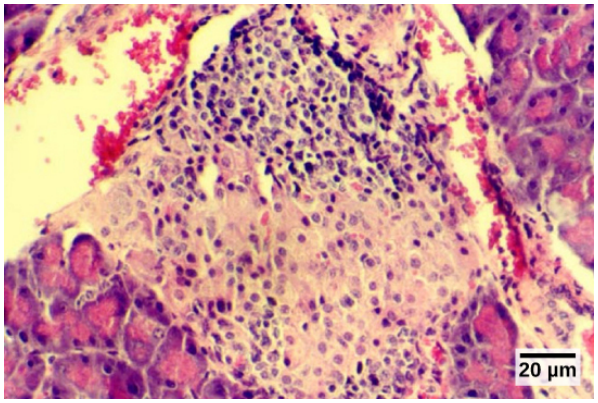


Figure 28.15 The islets of Langerhans are clusters of endocrine cells found in the pancreas; they stain lighter than surrounding cells. (credit: modification of work by Muhammad T. Tabiin, Christopher P. White, Grant Morahan, and Bernard E. Tuch; scale-bar data from Matt Russell)

Pineal Gland

The pineal gland produces melatonin. The rate of melatonin production is affected by the photoperiod (amount of light in a 24-hour period). Nerves from the visual pathways innervate the pineal gland. During the day (light), little melatonin is produced; however, melatonin production increases during the night (dark). In some mammals, melatonin has an inhibitory affect on reproductive functions by decreasing production and maturation of sperm, oocytes, and reproductive organs. Lastly, melatonin is involved in biological rhythms, particularly circadian rhythms such as the sleep-wake cycle and eating habits.

Gonads

The gonads—the male testes and female ovaries—produce steroid hormones. The testes produce androgens, testosterone being the most prominent, which allow for the development of secondary sex characteristics and the production of sperm cells. The ovaries produce estradiol and progesterone, which cause secondary sex characteristics and prepare the body for childbirth.

Endocrine Glands and their Associated Hormones

Endocrine Gland	Associated Hormones	Effect
Hypothalamus	releasing and inhibiting hormones	regulate hormone release from pituitary gland; produce oxytocin; produce uterine contractions and milk secretion in females
	antidiuretic hormone (ADH)	water reabsorption from kidneys; vasoconstriction to increase blood pressure
Pituitary (Anterior)	growth hormone (GH)	promotes growth of body tissues, protein synthesis; metabolic functions
	prolactin (PRL)	promotes milk production
	thyroid stimulating hormone (TSH)	stimulates thyroid hormone release
	adrenocorticotropic hormone (ACTH)	stimulates hormone release by adrenal cortex, glucocorticoids
	follicle-stimulating hormone (FSH)	stimulates gamete production (both ova and sperm); secretion of estradiol
	luteinizing hormone (LH)	stimulates androgen production by gonads; ovulation, secretion of progesterone

Table 28.1

Endocrine Glands and their Associated Hormones

Endocrine Gland	Associated Hormones	Effect
	melanocyte-stimulating hormone (MSH)	stimulates melanocytes of the skin increasing melanin pigment production.
Pituitary (Posterior)	antidiuretic hormone (ADH)	stimulates water reabsorption by kidneys
	oxytocin	stimulates uterine contractions during childbirth; milk ejection; stimulates ductus deferens and prostate gland contraction during emission
Thyroid	thyroxine, triiodothyronine	stimulate and maintain metabolism; growth and development
	calcitonin	reduces blood Ca^{2+} levels
Parathyroid	parathyroid hormone (PTH)	increases blood Ca^{2+} levels
Adrenal (Cortex)	aldosterone	increases blood Na^{+} levels; increase K^{+} secretion
	cortisol, corticosterone, cortisone	increase blood glucose levels; anti-inflammatory effects
Adrenal (Medulla)	epinephrine, norepinephrine	stimulate fight-or-flight response; increase blood gluclose levels; increase metabolic activities
Pancreas	insulin	reduces blood glucose levels
	glucagon	increases blood glucose levels
Pineal gland	melatonin	regulates some biological rhythms and protects CNS from free radicals
Testes	androgens	regulate, promote, increase or maintain sperm production; male secondary sexual characteristics
Ovaries	estrogen	promotes uterine lining growth; female secondary sexual characteristics
	progestins	promote and maintain uterine lining growth

Table 28.1

28.2 | How Animals Reproduce

Introduction

> “ Reproduction is so primitive and fundamental a function of vital organisms that the mechanism by which it is assured is highly complex and not yet clearly understood. It is not necessarily connected with sex, nor is sex necessarily connected with reproduction.”

Henry Havelock Ellis, in *Psychology of Sex*, 1933

Some animals produce offspring through asexual reproduction while other animals produce offspring through sexual reproduction. Both methods have advantages and disadvantages. **Asexual reproduction** produces offspring that are genetically identical to the parent because the offspring are all clones of the original parent. A single individual can produce offspring asexually and large numbers of offspring can be produced quickly; these are two advantages that asexually reproducing organisms have over sexually reproducing organisms. In a stable or predictable environment, asexual reproduction is an effective means of reproduction because all the offspring will be adapted to that environment. In an unstable or unpredictable environment, species that reproduce asexually may be at a disadvantage because all the offspring are genetically identical and may not be adapted to different conditions.

During **sexual reproduction**, the genetic material of two individuals is combined to produce genetically diverse offspring that differ from their parents. The genetic diversity of sexually produced offspring is thought to give sexually reproducing individuals greater fitness because more of their offspring may survive and reproduce in an unpredictable or changing environment. Species that reproduce sexually (and have separate sexes) must maintain two different types of individuals, males and females. Only half the population (females) can produce the offspring, so fewer offspring will be produced when compared to asexual reproduction. This is a disadvantage of sexual reproduction compared to asexual reproduction.

Asexual Reproduction

Asexual reproduction occurs in prokaryotic microorganisms (bacteria and archaea) and in many eukaryotic, single-celled and multi-celled organisms, both plants and animals. There are several ways that animals reproduce asexually, the details of which vary among individual species.

Fission

Fission, also called binary fission, occurs in some invertebrate, multi-celled organisms. It is in some ways analogous to the process of binary fission of single-celled prokaryotic organisms. The term fission is applied to instances in which an organism appears to split itself into two parts and, if necessary, regenerate the missing parts of each new organism. For example, some flatworms, such as *Dugesia dorotocephala*, are able to separate their bodies into head and tail regions and then regenerate the missing half in each of the two new organisms. Sea anemones (Cnidaria), such as species of the genus *Anthopleura* (**Figure 28.16**), will divide along the oral-aboral axis, and sea cucumbers (Echinodermata) of the genus *Holothuria*, will divide into two halves across the oral-aboral axis and regenerate the other half in each of the resulting individuals.

Figure 28.16 The *Anthopleura artemisia* sea anemone can reproduce through fission.

Budding

Budding is a form of asexual reproduction that results from the outgrowth of a part of the body leading to a separation of the “bud” from the original organism and the formation of two individuals, one smaller than the other. Budding occurs commonly in some invertebrate animals such as hydras and corals. In hydras, a bud forms that develops into an adult and breaks away from the main body (**Figure 28.17**).

(a) (b)

Figure 28.17 (a) Hydra reproduce asexually through budding: a bud forms on the tubular body of an adult hydra, develops a mouth and tentacles, and then detaches from its parent. The new hydra is fully developed and will find its own location for attachment. (b) Some coral, such as the *Lophelia pertusa* shown here, can reproduce through budding. (credit b: modification of work by Ed Bowlby, NOAA/Olympic Coast NMS; NOAA/OAR/Office of Ocean Exploration)

Fragmentation

Fragmentation is the breaking of an individual into parts followed by regeneration. If the animal is capable of fragmentation, and the parts are big enough, a separate individual will regrow from each part. Fragmentation may occur through accidental damage, damage from predators, or as a natural form of reproduction. Reproduction through fragmentation is observed in sponges, some cnidarians, turbellarians, echinoderms, and annelids. In some sea stars, a new individual can be regenerated from a broken arm and a piece of the central disc. This sea star (**Figure 28.18**) is in the process of growing a complete sea star from an arm that has been cut off. Fisheries workers have been known to try to kill the sea stars eating their clam or oyster beds by cutting them in half and throwing them back into the ocean. Unfortunately for the workers, the two parts can each regenerate a new half, resulting in twice as many sea stars to prey upon the oysters and clams.

(a) (b)

Figure 28.18 (a) *Linckia multifora* is a species of sea star that can reproduce asexually via fragmentation. In this process, (b) an arm that has been shed grows into a new sea star. (credit a: modifiction of work by Dwayne Meadows, NOAA/NMFS/OPR)

Parthenogenesis

Parthenogenesis is a form of asexual reproduction in which an egg develops into an individual without being fertilized. The resulting offspring can be either haploid or diploid, depending on the process and the particular species. Parthenogenesis occurs in invertebrates such as water fleas, rotifers, aphids, stick insects, and ants, wasps, and bees. Ants, bees, and wasps use parthenogenesis to produce haploid males (drones). The diploid females (workers and queens) are the result of a fertilized egg.

Some vertebrate animals—including some reptiles, amphibians, and fish—also reproduce through parthenogenesis. Parthenogenesis has been observed in species in which the sexes were separated in terrestrial or marine zoos. Two female Komodo dragons, a hammerhead shark, and a blacktop shark have produced parthenogenic young, even when the females have been isolated from males. It is possible that these instances of parthenogenesis occurred in response to unusual circumstances stemming from captivity, and would normally not occur.

Sexual Reproduction

Sexual reproduction is the combination of reproductive cells from two individuals, each contributing a haploid gamete, to generate genetically unique offspring. The nature of the individuals that produce the two kinds of gametes can vary, having for example separate sexes or both sexes in each individual.

Hermaphroditism

Hermaphroditism occurs in animals in which one individual has both male and female reproductive systems. Invertebrates such as earthworms, slugs, tapeworms, and snails (**Figure 28.19**) are often hermaphroditic. Hermaphrodites may self-fertilize, but typically they will mate with another of their species, fertilizing each other and both producing offspring. Self-fertilization is more common in animals that have limited mobility or are not motile, such as barnacles and clams. Many species have specific mechanisms in place to prevent self-fertilization, because it is an extreme form of inbreeding and usually produces less fit offspring.

(a) (b)

Figure 28.19 Many (a) snails are hermaphrodites. When two individuals (b) mate, they can produce up to 100 eggs each. (credit a: modification of work by Assaf Shtilman; credit b: modification of work by "Schristia"/Flickr)

Fertilization

The fusion of a sperm and an egg is a process called fertilization. This can occur either inside (internal fertilization) or outside (external fertilization) the body of the female. Humans provide an example of the former, whereas frog reproduction is an example of the latter.

External Fertilization

External fertilization usually occurs in aquatic environments where both eggs and sperm are released into the water. After the sperm reaches the egg, fertilization takes place. Most external fertilization happens during the process of spawning where one or several females release their eggs and the male(s) release sperm in the same area, at the same time. The spawning may be triggered by environmental signals, such as water temperature or the length of daylight. Nearly all fish spawn, as do crustaceans (such as crabs and shrimp), mollusks (such as oysters), squid, and echinoderms (such as sea urchins and sea cucumbers). Frogs, corals, mayflies, and mosquitoes also spawn (**Figure 28.20**).

Figure 28.20 During sexual reproduction in toads, the male grasps the female from behind and externally fertilizes the eggs as they are deposited. (credit: Bernie Kohl)

Internal Fertilization

Internal fertilization occurs most often in terrestrial animals, although some aquatic animals also use this method. Internal fertilization may occur by the male directly depositing sperm in the female during mating. It may also occur by the male depositing sperm in the environment, usually in a protective structure, which a female picks up to deposit the sperm in her reproductive tract. There are three ways that offspring are produced following internal fertilization. In oviparity, fertilized eggs are laid outside the female's body and develop there, receiving nourishment from the yolk that is a part of the egg (**Figure 28.21a**). This occurs in insects, some bony fish, some reptiles, a few cartilaginous fish, some amphibians, a few mammals, and all birds. Most non-avian reptiles and insects produce leathery eggs, while birds and some turtles produce eggs with high concentrations of calcium carbonate in the shell, making them hard. Chicken eggs are an example of a hard shell. The eggs of the egg-laying mammals such as the platypus and echidna are leathery.

In ovoviparity, fertilized eggs are retained in the female, and the embryo obtains its nourishment from the egg's yolk. The eggs are retained in the female's body until they hatch inside of her, or she lays the eggs right before they hatch. This process helps protect the eggs until hatching. This occurs in some bony fish (like the platyfish *Xiphophorus maculatus*, **Figure 28.21b**), some sharks, lizards, some snakes (garter snake *Thamnophis sirtalis*), some vipers, and some invertebrate animals (Madagascar hissing cockroach *Gromphadorhina portentosa*).

In viviparity the young are born alive. They obtain their nourishment from the female and are born in varying states of maturity. This occurs in most mammals (**Figure 28.21c**), some cartilaginous fish, and a few reptiles.

(a) (b) (c)

Figure 28.21 In (a) oviparity, young develop in eggs outside the female body, as with these *Harmonia axydridis* beetles hatching. Some aquatic animals, like this (b) pregnant *Xiphophorus maculatus* are ovoviparous, with the egg developing inside the female and nutrition supplied primarily from the yolk. In mammals, nutrition is supported by the placenta, as was the case with this (c) newborn squirrel. (credit b: modification of work by Gourami Watcher; credit c: modification of work by "audreyjm529"/Flickr)

Hormonal Control of Reproduction

In the animal kingdom there are many interesting examples of reproductive strategies, and hormones are involved in regulating all of those. Rather than attempt to cover the wide diversity of strategies and hormones, we will concentrate on the hormonal control of reproduction in a single species, *Homo sapiens*.

The human male and female reproductive cycles are controlled by the interaction of hormones from the hypothalamus and anterior pituitary with hormones from reproductive tissues and organs. In both sexes, the hypothalamus monitors and causes the release of hormones from the anterior pituitary gland. When the reproductive hormone is required, the hypothalamus sends a **gonadotropin-releasing hormone (GnRH)** to the anterior pituitary. This causes the release of **follicle stimulating hormone (FSH)** and **luteinizing hormone (LH)** from the anterior pituitary into the blood. Although these hormones are named after their functions in female reproduction, they are produced in both sexes and play important roles in controlling reproduction. Other hormones have specific functions in the male and female reproductive systems.

Male Hormones

At the onset of puberty, the hypothalamus causes the release of FSH and LH into the male system for the first time. FSH enters the testes and stimulates the Sertoli cells located in the walls of the seminiferous tubules to begin promoting spermatogenesis (**Figure 28.22**). LH also enters the testes and stimulates the interstitial cells of Leydig, located in between the walls of the seminiferous tubules, to make and release testosterone into the testes and the blood.

Testosterone stimulates spermatogenesis. This hormone is also responsible for the secondary sexual characteristics that develop in the male during adolescence. The secondary sex characteristics in males include a deepening of the voice, the growth of facial, axillary, and pubic hair, an increase in muscle bulk, and the beginnings of the sex drive.

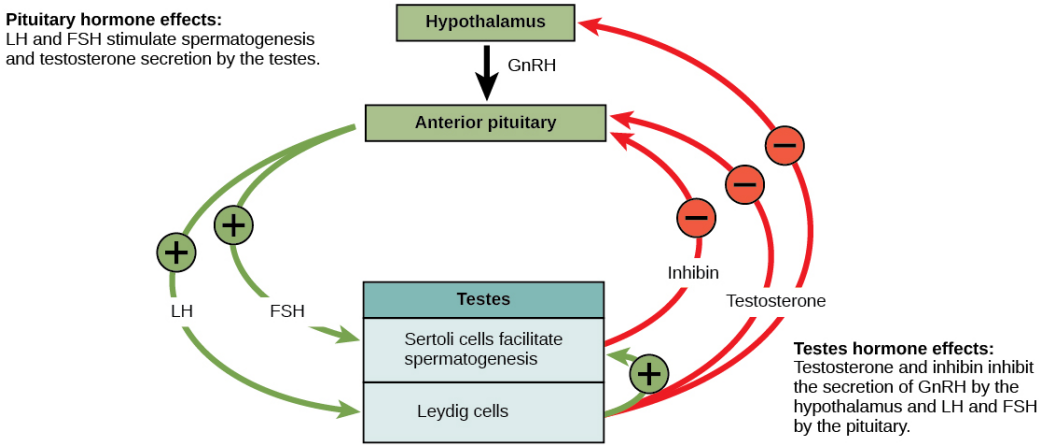


Figure 28.22 Hormones control sperm production in a negative feedback system.

A negative feedback system occurs in the male with rising levels of testosterone acting on the hypothalamus and anterior pituitary to inhibit the release of GnRH, FSH, and LH. In addition, the Sertoli cells produce the hormone inhibin, which is released into the blood when the sperm count is too high. This inhibits the release of GnRH and FSH, which will cause spermatogenesis to slow down. If the sperm count reaches a low of 20 million/mL, the Sertoli cells cease the release of inhibin, and the sperm count increases.

Female Hormones

The control of reproduction in females is more complex. The female reproductive cycle is divided into the ovarian cycle and the menstrual cycle. The ovarian cycle governs the preparation of endocrine tissues and release of eggs, while the menstrual cycle governs the preparation and maintenance of the uterine lining (**Figure 28.23**). These cycles are coordinated over a 22–32 day cycle, with an average length of 28 days.

As with the male, the GnRH from the hypothalamus causes the release of the hormones FSH and LH from the anterior pituitary. In addition, estrogen and progesterone are released from the developing follicles. As with testosterone in males, estrogen is responsible for the secondary sexual characteristics of females. These include breast development, flaring of the hips, and a shorter period for bone growth.

The Ovarian Cycle and the Menstrual Cycle

The ovarian and menstrual cycles are regulated by hormones of the hypothalamus, pituitary, and ovaries (**Figure 28.23**). The ebb and flow of the hormones causes the ovarian and menstrual cycles to advance. The ovarian and menstrual cycles occur concurrently. The first half of the ovarian cycle is the follicular phase. Slowly rising levels of FSH cause the growth of follicles on the surface of the ovary. This process prepares the egg for ovulation. As the follicles grow, they begin releasing estrogen. The first few days of this cycle coincide with menstruation or the sloughing off of the functional layer of the endometrium in the uterus. After about five days, estrogen levels rise and the menstrual cycle enters the proliferative phase. The endometrium begins to regrow, replacing the blood vessels and glands that deteriorated during the end of the last cycle.

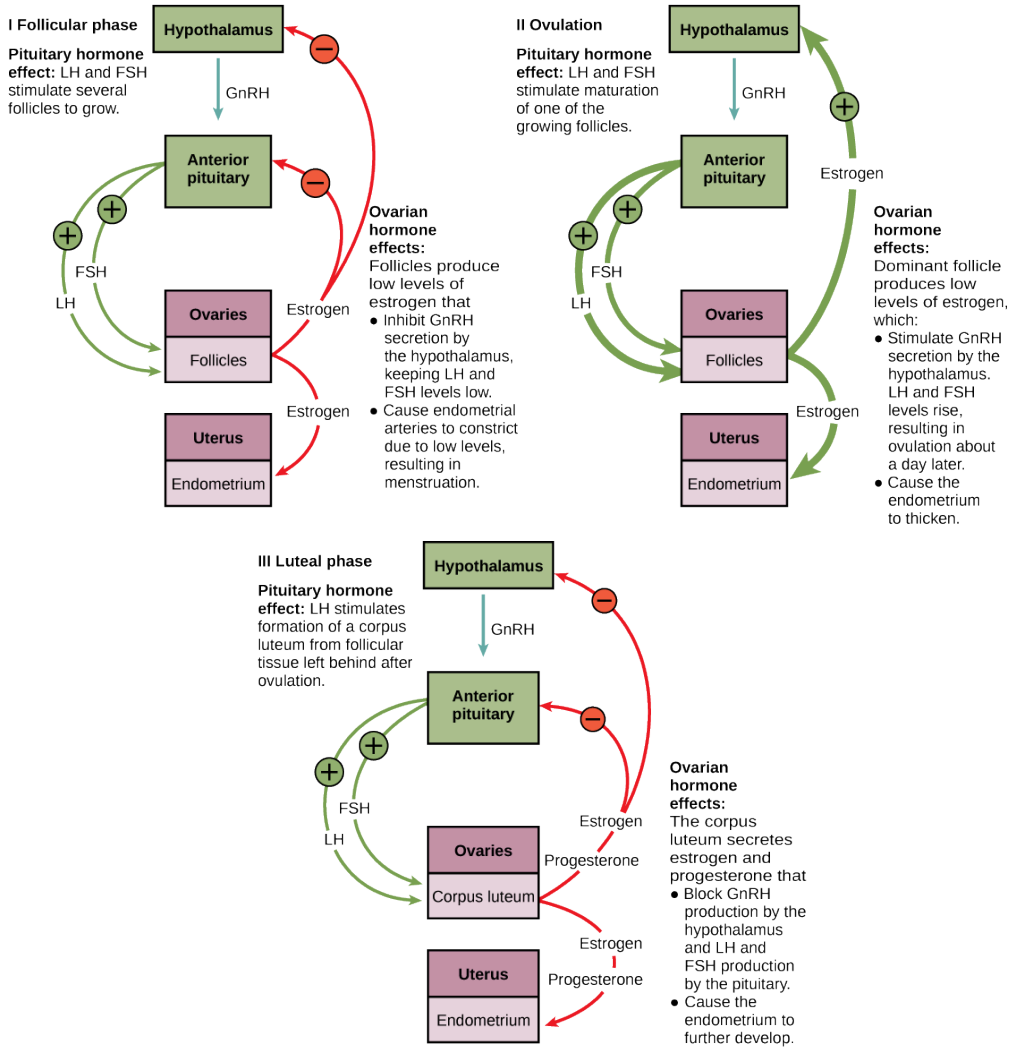


Figure 28.23 The ovarian and menstrual cycles of female reproduction are regulated by hormones produced by the hypothalamus, pituitary, and ovaries.

Just prior to the middle of the cycle (approximately day 14), the high level of estrogen causes FSH and especially LH to rise rapidly then fall. The spike in LH causes the most mature follicle to rupture and release its egg. This is ovulation. The follicles that did not rupture degenerate and their eggs are lost. The level of estrogen decreases when the extra follicles degenerate.

Following ovulation, the ovarian cycle enters its luteal phase and the menstrual cycle enters its secretory phase, both of which run from about day 15 to 28. The luteal and secretory phases refer to changes in the ruptured follicle. The cells in the follicle undergo physical changes and produce a structure called a corpus luteum. The corpus luteum produces estrogen and progesterone. The progesterone facilitates the regrowth of the uterine lining and inhibits the release of further FSH and LH. The uterus is being prepared to accept a fertilized egg, should it occur during this cycle. The inhibition of FSH and LH prevents any further eggs and follicles from developing, while the progesterone is elevated. The level of estrogen produced by the corpus luteum increases to a steady level for the next few days.

If no fertilized egg is implanted into the uterus, the corpus luteum degenerates and the levels of estrogen and progesterone decrease. The endometrium begins to degenerate as the progesterone levels drop, initiating the next menstrual cycle. The decrease in progesterone also allows the hypothalamus to send GnRH to the anterior pituitary, releasing FSH and LH and starting the cycles again.

Gestation

Pregnancy begins with the fertilization and implantation of an egg and continues through to the birth of the individual. The length of time of gestation, or the gestation period, in humans is 266 days and is similar in other great apes. Gestation periods in other animals range from 12-13 days in the American opossum to the 660 day gestation period of the African

elephant.

Within 24 hours of fertilization, the egg nucleus has finished meiosis and the egg and sperm nuclei fuse. With fusion, the cell is known as a zygote. The zygote initiates cleavage and the developing embryo travels through the oviduct to the uterus. The developing embryo must implant into the wall of the uterus within seven days, or it will deteriorate and die. The outer layers of the developing embryo or blastocyst grow into the endometrium by digesting the endometrial cells, and healing of the endometrium closes up the blastocyst into the tissue. Another layer of the blastocyst, the chorion, begins releasing a hormone called **human beta chorionic gonadotropin (β-HCG)**, which makes its way to the corpus luteum and keeps that structure active. This ensures adequate levels of progesterone that will maintain the endometrium of the uterus for the support of the developing embryo. Pregnancy tests determine the level of β-HCG in urine or serum. If the hormone is present, the test is positive.

The gestation period is divided into three equal periods or trimesters. During the first two-to-four weeks of the first trimester, nutrition and waste are handled by the endometrial lining through diffusion. As the trimester progresses, the outer layer of the embryo begins to merge with the endometrium, and the placenta forms. The placenta takes over the nutrient and waste requirements of the embryo and fetus, with the mother's blood passing nutrients to the placenta and removing waste from it. Chemicals from the fetus, such as bilirubin, are processed by the mother's liver for elimination. Some of the mother's immunoglobulins will pass through the placenta, providing passive immunity against some potential infections.

Internal organs and body structures begin to develop during the first trimester. By five weeks, limb buds, eyes, the heart, and liver have been basically formed. By eight weeks, the term fetus applies, and the body is essentially formed (**Figure 28.24a**). The individual is about five centimeters (two inches) in length and many of the organs, such as the lungs and liver, are not yet functioning. Exposure to any toxins is especially dangerous during the first trimester, as all of the body's organs and structures are going through initial development. Anything that interferes with chemical signaling during that development can have a severe effect on the fetus' survival.

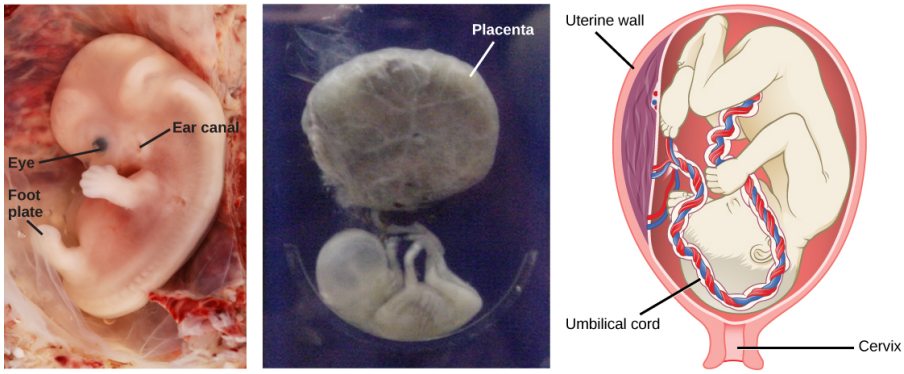


Figure 28.24 (a) Fetal development is shown at nine weeks gestation. (b) This fetus is just entering the second trimester, when the placenta takes over more of the functions performed as the baby develops. (c) There is rapid fetal growth during the third trimester. (credit a: modification of work by Ed Uthman; credit b: modification of work by National Museum of Health and Medicine; credit c: modification of work by Gray's Anatomy)

During the second trimester, the fetus grows to about 30 cm (about 12 inches) (**Figure 28.24b**). It becomes active and the mother usually feels the first movements. All organs and structures continue to develop. The placenta has taken over the functions of nutrition and waste elimination and the production of estrogen and progesterone from the corpus luteum, which has degenerated. The placenta will continue functioning up through the delivery of the baby. During the third trimester, the fetus grows to 3 to 4 kg (6.5–8.5 lbs.) and about 50 cm (19–20 inches) long (**Figure 28.24c**). This is the period of the most rapid growth during the pregnancy as all organ systems continue to grow and develop.

Labor is the name given to the muscular contractions that expel the fetus and placenta from the uterus. Toward the end of the third trimester, estrogen causes receptors on the uterine wall to develop and bind the hormone oxytocin. At this time, the fetus usually reorients, facing forward and down with the back or crown of the head pushing on the cervix (uterine opening). This causes the cervix to stretch and nerve impulses are sent to the hypothalamus, which signals the release of oxytocin from the posterior pituitary. Oxytocin causes smooth muscle in the uterine wall to contract. At the same time, the placenta releases prostaglandins into the uterus, increasing the contractions. A positive feedback relay occurs between the uterus, hypothalamus, and the posterior pituitary to assure an adequate supply of oxytocin. As more smooth muscle cells are recruited, the contractions increase in intensity and force. As noted previously, this is a good example of a positive feedback loop – the stimulus causes the production of a hormone that increases the stimulus.

There are three stages to labor. During stage one, the cervix thins and dilates. This is necessary for the baby and placenta to

be expelled during birth. The cervix will eventually dilate to about 10 cm. During stage two, the baby is expelled from the uterus. The uterus contracts and the mother pushes as she compresses her abdominal muscles to aid the delivery. The last stage is the passage of the placenta after the baby has been born and the organ has completely disengaged from the uterine wall. If labor should stop before stage two is reached, synthetic oxytocin, known as Pitocin, can be administered to restart and maintain labor.

28.3 | Urinary System

Introduction

> “ What is man, when you come to think upon him, but a minutely set, ingenious machine for turning, with infinite artfulness, the red wine of Shiraz into urine?”

Baroness Karen Blixen, in *The Dreamers*, 1943

Osmoregulation and Osmotic Balance

Osmosis is the diffusion of water across a membrane in response to osmotic pressure caused by differences in the solute molecules on either side of the membrane. **Osmoregulation** is the active homeostatic process of the water content of an organism, involving movement of solutes across membranes so that water moves in response to the ion concentration . It might be beneficial to review osmosis here - **the section called “Osmosis”**. Osmo regulation involves control of the water and solute content of all the fluids in the animal body. There are three general fluid pools in the typical animal: the blood plasma, the cytosol within cells, and the interstitial fluid (the fluid that exists in the spaces between cells and tissues of the body). See **Figure 28.25** for a review of how solute concentrations affect the movement of water across plasma membranes.

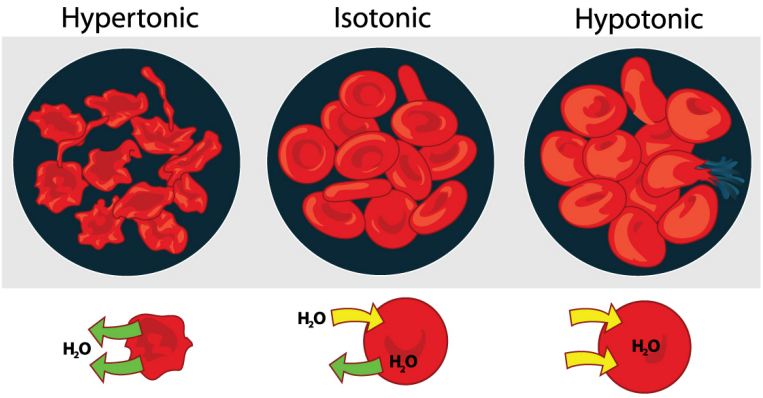


Figure 28.25 Cells placed in a hypertonic environment tend to shrink due to loss of water. In a hypotonic environment, cells tend to swell due to intake of water. The blood maintains an isotonic environment so that cells neither shrink nor swell. (credit: Mariana Ruiz Villareal)

Need for Osmoregulation

The body does not exist in isolation. There is a constant input of water and electrolytes into the system; osmoregulation is thus a constant process. 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 either gain or lose 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.

Transport of Electrolytes across Plasmal Membranes

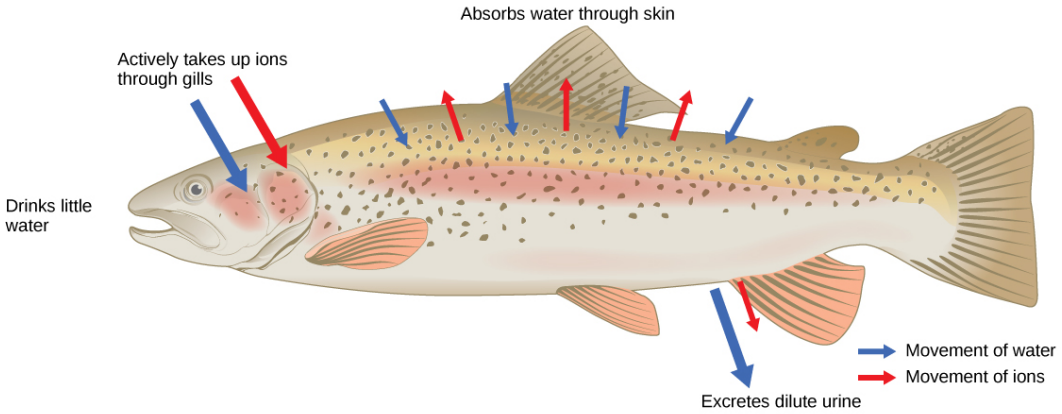
Electrolytes, such as sodium chloride, ionize in water, meaning that they dissociate into their component ions. In water, sodium chloride (NaCl), dissociates into the sodium ion (Na^+) and the chloride ion ($Cl^–$). The most important ions, whose concentrations are very closely regulated in body fluids, are the cations sodium (Na^+), potassium (K^+), calcium (Ca^{+2}), magnesium (Mg^{+2}), and the anions chloride (Cl^-), carbonate (CO_3^{-2}), bicarbonate (HCO_3^-), and phosphate(PO_3^-). Electrolytes are lost from the body during urination and perspiration. For this reason, athletes are encouraged to replace electrolytes and fluids during periods of increased activity and perspiration.

Osmotic pressure is influenced by the concentration of solutes in a solution. It is directly proportional to the concentration of solute atoms or molecules, and not dependent on the size of the solute molecules. Because some compounds (known as electrolytes) dissociate into their component ions, they 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.

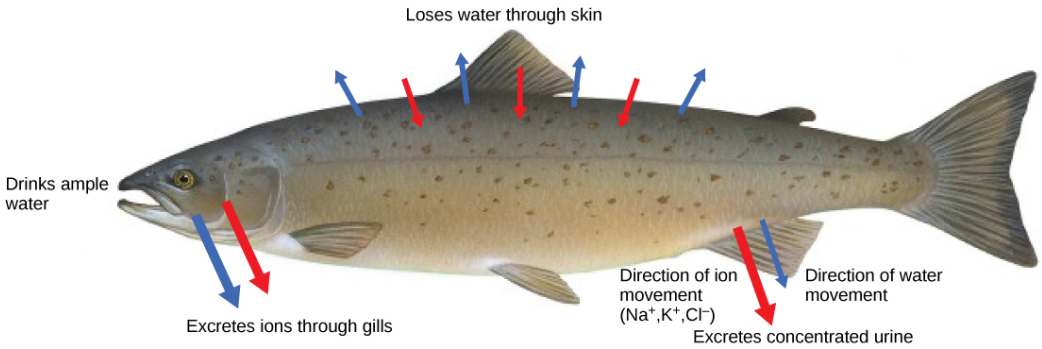
Water can pass through membranes by passive diffusion. If electrolyte ions could passively diffuse across membranes, it would be impossible to maintain specific concentrations of ions in each fluid compartment therefore they require special mechanisms to cross the semi-permeable membranes in the body. 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 fresh water and part in sea water. Organisms like the salmon 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 28.26a**. 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 28.26b**. Most marine invertebrates, on the other hand, may be isotonic with sea water; these are known as **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 sea water. 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 have a rectal gland which secretes salt and assists in osmoregulation.



(a) Osmoregulation in a freshwater environment



(b) Osmoregulation in a saltwater environment

Figure 28.26 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)

Kidneys and Osmoregulatory Organs

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 28.27**, 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. 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. The filtrate coming out of the kidneys is called **urine**.

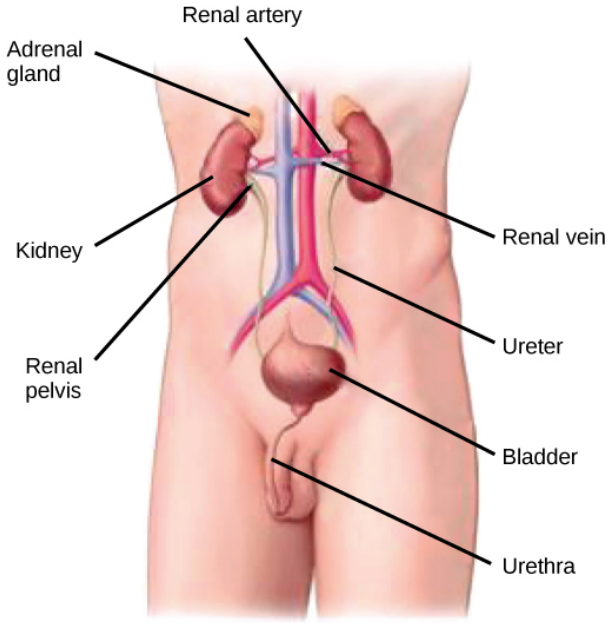


Figure 28.27 Kidneys filter the blood, producing urine that is stored in the bladder prior to elimination through the urethra. (credit: modification of work by NCI)

Kidney Structure

Externally, the kidneys are surrounded by three layers, illustrated in **Figure 28.28**. 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.

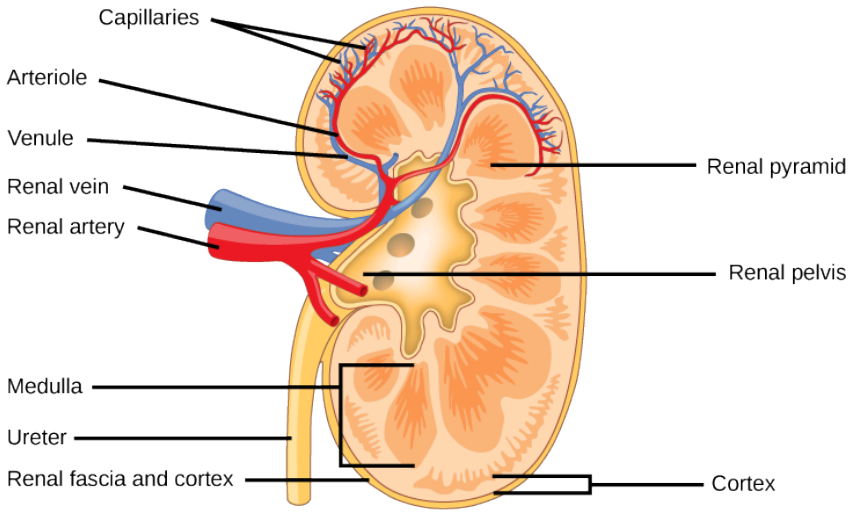


Figure 28.28 The internal structure of the kidney is shown. (credit: modification of work by NCI)

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 and ends with the exiting of the renal veins to join the inferior vena cava.

As mentioned previously, the functional unit of the kidney is the nephron, illustrated in Figure 28.29. Each kidney is made up of over one million nephrons that dot the renal cortex, giving it a granular appearance when sectioned sagittally. A nephron consists of three parts—a renal corpuscle, a renal tubule, and the associated capillary network, which originates from the arteries that supply blood to the kidney.

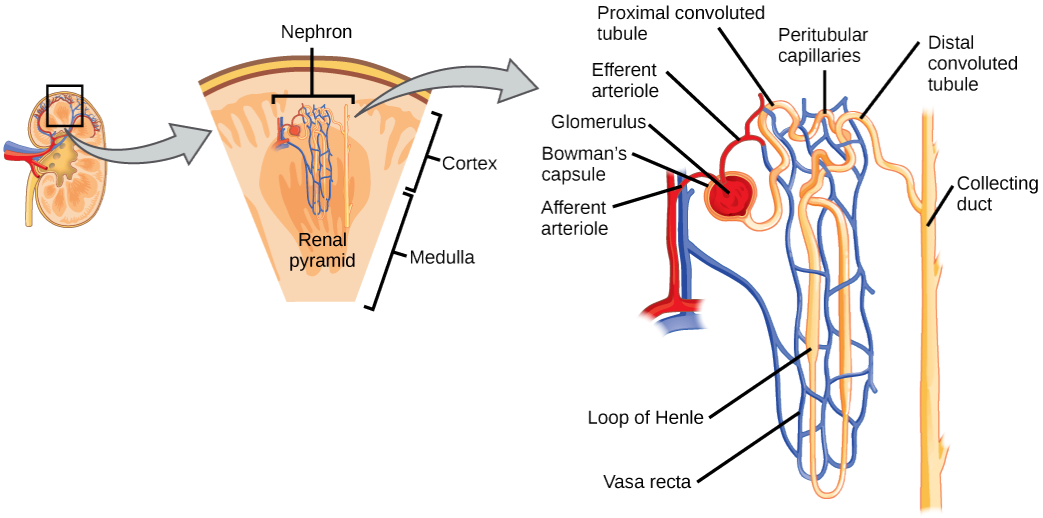


Figure 28.29 The nephron is the functional unit of the kidney. The glomerulus and convoluted tubules are located in the kidney cortex, while collecting ducts are located in the pyramids of the medulla. (credit: modification of work by NIDDK)

Renal Corpuscle

The renal corpuscle, located in the renal cortex, is made up of a network of capillaries known as the **glomerulus** and the capsule, a cup-shaped chamber that surrounds it, called the glomerular or **Bowman's capsule**.

Renal Tubule

The renal tubule is a long and convoluted structure that emerges from the glomerulus and can be divided into three parts based on function. The first part is called the **proximal convoluted tubule (PCT)** due to its proximity to the glomerulus; it stays in the renal cortex. The second part is called the **loop of Henle**, or nephritic loop, because it forms a loop (with **descending** and **ascending limbs**) that goes through the renal medulla. The third part of the renal tubule is called the **distal convoluted tubule (DCT)** and this part is also restricted to the renal cortex. The DCT, which is the last part of the nephron, connects and empties its contents into collecting ducts that line the medullary pyramids. The collecting ducts amass contents from multiple nephrons and fuse together as they enter the papillae of the renal medulla.

Capillary Network within the Nephron

The capillary network that originates from the renal arteries supplies the nephron with blood that needs to be filtered. The branch that enters the glomerulus is called the afferent arteriole. The branch that exits the glomerulus is called the efferent arteriole. Within the glomerulus, the network of capillaries is called the glomerular capillary bed. Once the efferent arteriole exits the glomerulus, it forms the peritubular capillary network, which surrounds and interacts with parts of the renal 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**. The high arterial pressure and the permeable membranes of the glomerulus (see below) combine to accomplish this 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 to 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 28.30.

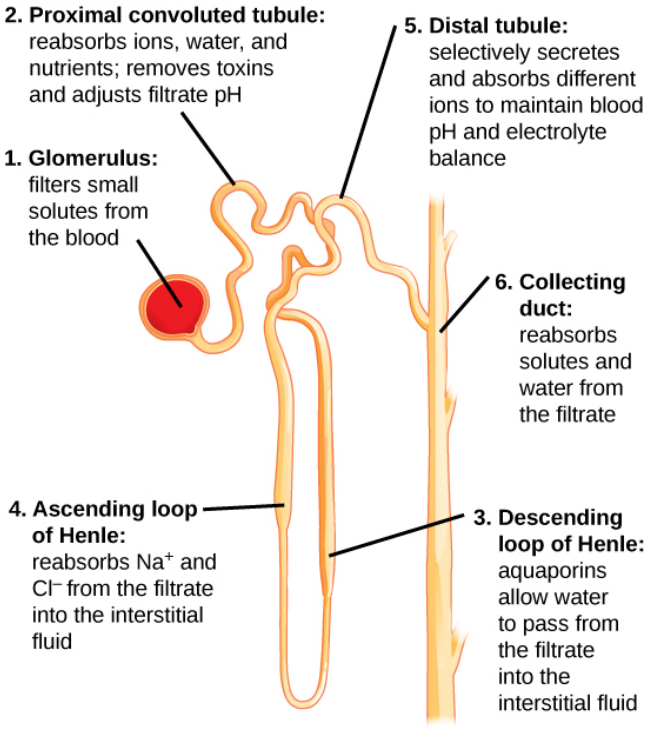


Figure 28.30 Each part of the nephron performs a different function in filtering waste and maintaining homeostatic balance. (1) The glomerulus forces small solutes out of the blood by pressure. (2) The proximal convoluted tubule reabsorbs ions, water, and nutrients from the filtrate into the interstitial 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 (NH_3) into the filtrate, where it reacts with H^+ to form NH_4^+. 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 HCO_3^- 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)

Glomerular Filtration

Glomerular filtration filters out most of the solutes due to high blood pressure and specialized membranes in the afferent arteriole. The blood pressure in the glomerulus is maintained independent of factors that affect systemic blood pressure. The "leaky" connections between the endothelial cells of the glomerular capillary network allow solutes to pass through easily. All solutes in the glomerular capillaries, except for macromolecules like proteins, pass through by passive diffusion. There is no energy requirement at this stage of the filtration process; high arterial blood pressure does the work at this stage.

Tubular Reabsorption and Secretion

Tubular reabsorption occurs in the PCT part of the renal tubule. Almost all nutrients (e.g. glucose, amino acids) 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 28.31**, it acts as a countercurrent multiplier. The vasa recta around it acts as the countercurrent exchanger.

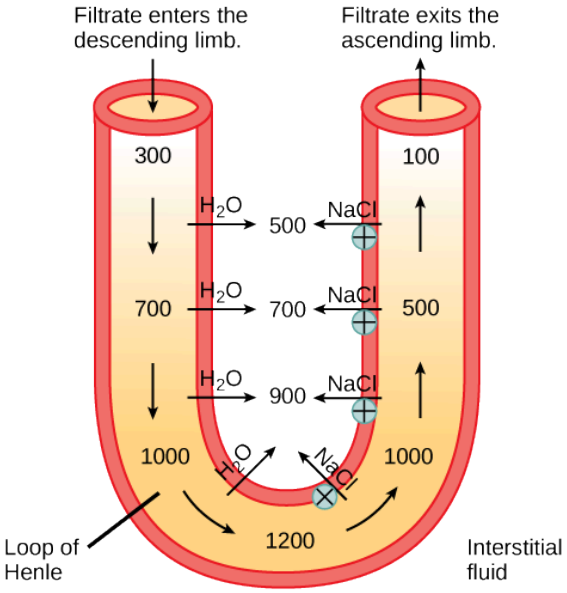


Figure 28.31 The loop of Henle acts as a countercurrent multiplier that uses energy to create concentration gradients. The descending limb is water permeable. Water flows from the filtrate to the interstitial fluid, so osmolality inside the limb increases as it descends into the renal medulla. At the 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 plasma 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).

Hypertension (high blood pressure) is a common problem for humans, and is usually treated with a variety of drugs that act on various processes occurring in the kidney. One class of hypertension drugs is the so-called "loop diuretics", which inhibit the reabsorption of Na^+ and Cl^- ions by the ascending limb of the loop of Henle. A side effect is that they increase urination. Why do you think this is the case?

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.

Nitrogenous Waste

Of the four major macromolecules in biological systems, both proteins and nucleic acids contain nitrogen. During the catabolism, or breakdown, of nitrogen-containing macromolecules, carbon, hydrogen, and oxygen are extracted and stored in the form of carbohydrates and fats. Excess nitrogen is excreted from the body. Nitrogenous wastes tend to form toxic **ammonia**, which raises the pH of body fluids. The formation of ammonia itself requires energy in the form of ATP and large quantities of water to dilute it out of a biological system. It is quite toxic even at relatively low concentrations. Animals that live in aquatic environments tend to release ammonia directly into the water in the urine; they have access to sufficient water to dilute this waste product to non-toxic levels. Animals that excrete ammonia are said to be ammonotelic. Terrestrial organisms have evolved other mechanisms to process and excrete nitrogenous wastes. The animals first detoxify ammonia by converting it into a relatively nontoxic form such as urea or uric acid. Mammals, including humans, produce urea,

whereas reptiles and many terrestrial invertebrates produce uric acid. Animals that secrete urea as the primary nitrogenous waste material are called ureotelic animals.

Nitrogenous Waste in Terrestrial Animals: Urea

Urea formation is the primary mechanism by which mammals convert ammonia to urea. Urea is made in the liver and excreted in urine. The overall chemical reaction by which ammonia is converted to urea is 2 NH_3 (ammonia) + CO_2 + 3 ATP + H_2O → H_2N-CO-NH_2 (urea) + 2 ADP + 4 P_i + AMP.

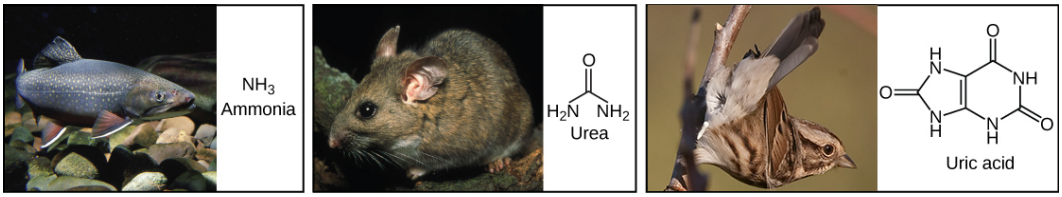
evolution CONNECTION

Excretion of Nitrogenous Waste

The theory of evolution proposes that life started in an aquatic environment. It is not surprising to see that biochemical pathways like the urea cycle evolved to adapt to a changing environment when terrestrial life forms evolved. Arid conditions probably led to the evolution of the uric acid pathway as a means of conserving water.

Nitrogenous Waste in Birds and Reptiles: Uric Acid

Birds, reptiles, and most terrestrial arthropods convert toxic ammonia to **uric acid** or the closely related compound guanine (guano) instead of urea. Mammals also form some uric acid during breakdown of nucleic acids. Uric acid is a compound similar to purines found in nucleic acids. It is water insoluble and tends to form a white paste or powder; it is excreted by birds, insects, and reptiles. Conversion of ammonia to uric acid requires more energy and is much more complex than conversion of ammonia to urea **Figure 28.32**, but the pay off is that uric acid requires much less water when excreted.



(a) Many invertebrates and aquatic species excrete ammonia.

(b) Mammals, many adult amphibians, and some marine species excrete urea.

(c) Insects, land snails, birds, and many reptiles excrete uric acid.

Figure 28.32 Nitrogenous waste is excreted in different forms by different species. These include (a) ammonia, (b) urea, and (c) uric acid. (credit a: modification of work by Eric Engbretson, USFWS; credit b: modification of work by B. "Moose" Peterson, USFWS; credit c: David A. Rintoul)

Gout

In some animals, uric acid can build up under certain conditions, or as consequence of a diet high in nitrogenous compounds (e.g. nucleotides). In those situations, uric acid tends to crystallize and form kidney stones. Uric acid buildup may also cause a painful condition called gout, where uric acid crystals accumulate in the joints, as illustrated in **Figure 28.33**. Food choices that reduce the amount of nitrogenous compounds in the diet help reduce the risk of gout. For example, tea, coffee, and chocolate have purine-like compounds, called xanthines, and should be avoided by people with gout and kidney stones.

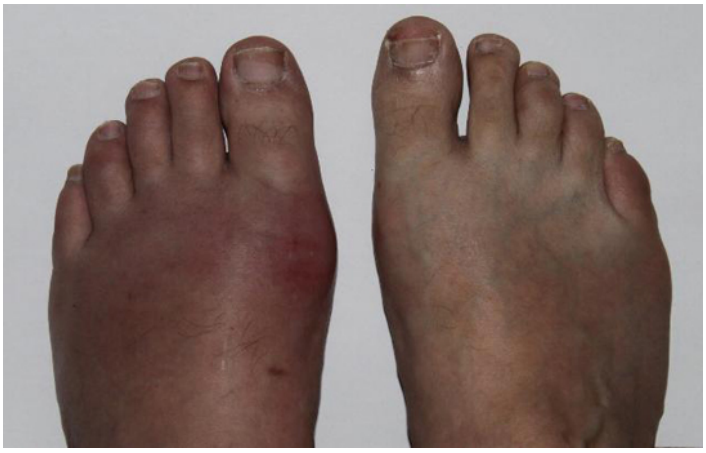


Figure 28.33 Gout causes the inflammation visible in this person's left big toe joint. (credit: "Gonzosft"/Wikimedia Commons)

29 | NERVOUS, SENSORY AND MUSCULOSKELETAL SYSTEMS

29.1 | Nervous System

Introduction

> "We can trace the development of a nervous system, and correlate with it the parallel phenomena of sensation and thought. We see with undoubting certainty that they go hand-in-hand. But we try to soar in a vacuum the moment we seek to comprehend the connexion between them... Man the object is separated by an impassable gulf from man the subject."

John Tyndall, British physicist, in *Fragments of Science for Unscientific People: A Series of Detached Essays, Lectures and Reviews*, 1892

The distinction between the brain and the mind, as described by Tyndall, is but one of many questions that have fascinated scientists regarding the human nervous system. Several Nobel Prizes have been awarded to scientists who have helped elucidate the workings of nerves and nervous systems, usually with the aid of studies in non-human organisms.

While you're reading this, 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 studying to finding a mate.

Diversity of Nervous Systems

Nervous systems throughout the animal kingdom vary in structure and complexity, as illustrated by the variety of animals shown in **Figure 29.1**. 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.

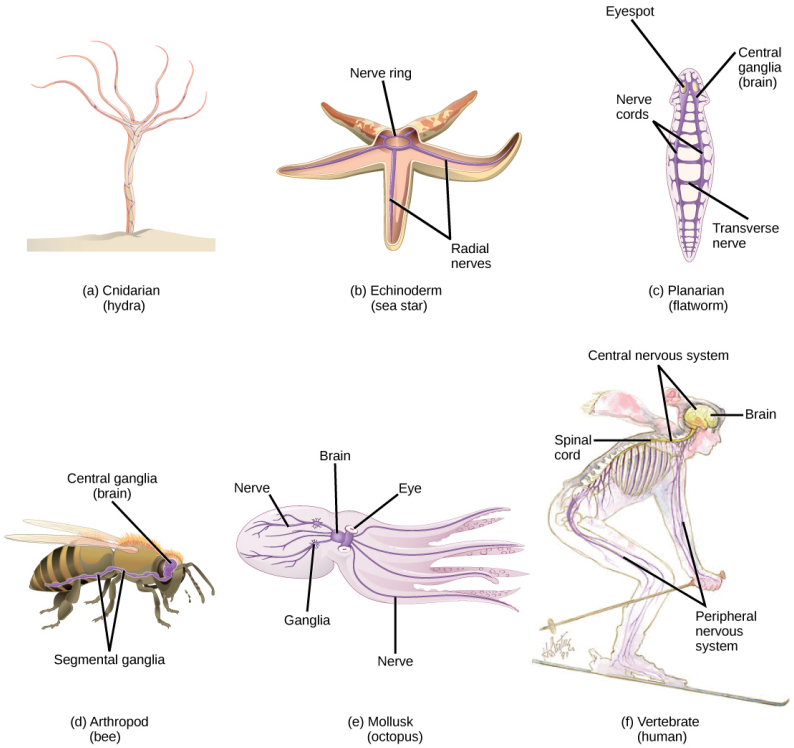


Figure 29.1 Nervous systems vary in structure and complexity. In (a) cnidarians, nerve cells form a decentralized nerve net. In (b) echinoderms, nerve cells are bundled into fibers called nerves. In animals exhibiting bilateral symmetry such as (c) planarians, neurons cluster into an anterior brain that 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.

Neurons and Glial Cells

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. This analogy might be oversimplified, however; recent evidence suggests that glial cells also usurp some of the signaling functions of 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.

There is great diversity in the types of neurons and glia that are present in different parts of the nervous system. There are three major functional types of neurons (and many different morphological types), and they share several important cellular components. But neurons are also highly specialized—different types of neurons have different sizes and shapes that relate to their functional roles. There are also several types of glial cells (astroglia, oligodendrocytes, Schwann cells, etc.) each

with different functions.

Neurons

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 29.2** 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.

Once a signal is received by the dendrite, it then travels 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 (known as neurotransmitters) 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 (a product of the glial cells), which acts as an insulator and greatly increases the speed of conduction. 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.

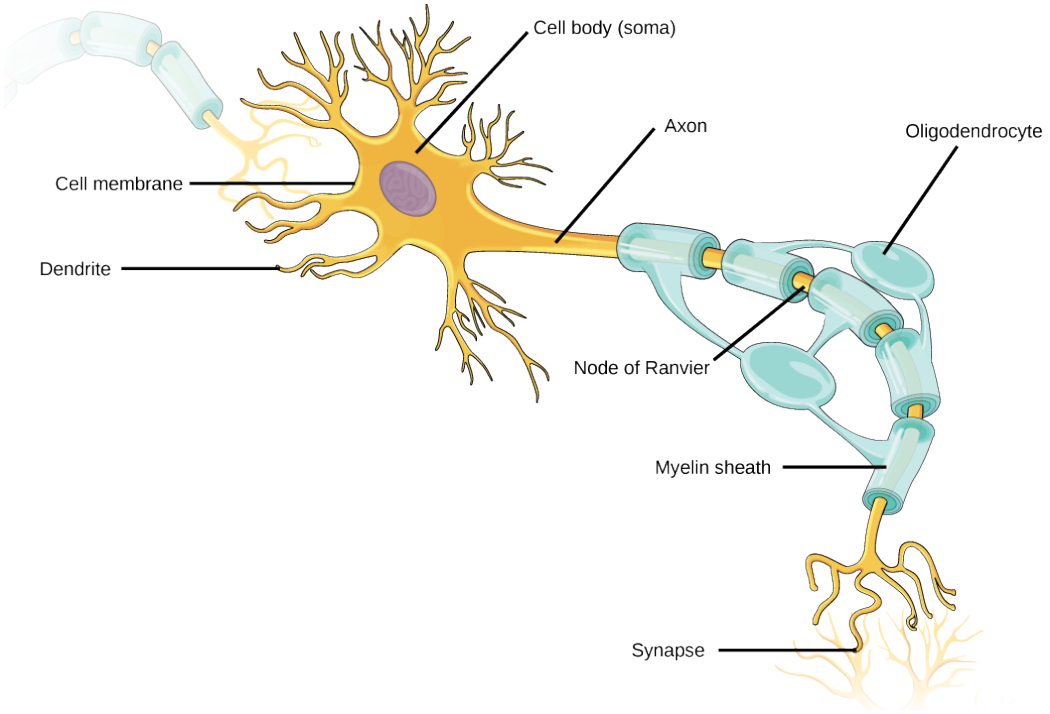


Figure 29.2 Neurons contain organelles common to many other cells, such as a nucleus and mitochondria. They also have more specialized structures, including dendrites and axons.

Types of Neurons

There are different types of neurons, and the functional role of a given neuron is intimately dependent on its structure. Although there are only three functional types of neurons **Figure 29.3**, an amazing diversity of neuron shapes and sizes can found in different parts of the nervous system (and across species).

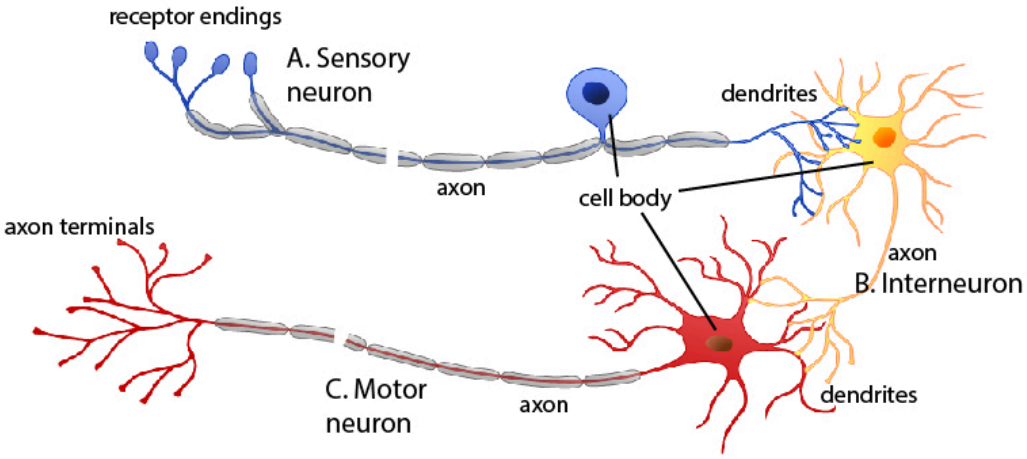


Figure 29.3 Neuron Types The three general classes of neurons; all have an input zone (recpetor endings, dendrites and/or the cell body), an axon, a cell body, and an output zone (axon terminals). A. Sensory neurons have receptor endings at one end that are sensitive to various stimuli (e.g. heat, pressure, light, etc.), a relatively long axon, and axon terminals that form synapses with dendrites at the other end. B. interneurons receive signals from sensory neurons via their dendrites at one end, have a relatively short axon, and pass signals to another neuron via axon terminals at the other end. C. Motor neurons receive signals via dendrites at one end, have a long axon, and transmit signals to muscles or glands at the other end. (Image by Eva Horne)

While there are many defined neuron cell shapes, neurons are broadly divided into three basic types: sensory, interneuron, and motor neuron. In general, **sensory neurons** detect information, either from the external environment or from internal sources. Examples of sensory neurons include the pain receptors in your skin and the photoreceptors in your retina. When activated by the signal to which they are attuned, they send information (via an action potential) to an interneuron. **Interneurons** both receive signals from other neurons and transmit signals to other neurons. The majority of the cells in your brain and spinal cord are interneurons, communicating only with other neurons. Interneurons can also send a signal to **motor neurons**, which control muscles and endocrine glands.

Glia

While glia are often thought of as the supporting cast of the nervous system, the number of glial cells in the human 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, contribute to the formation of cerebrospinal fluid, 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.

How Neurons Communicate

All functions performed by the nervous system—from a simple motor reflex to more advanced functions like making a memory or a decision—require neurons to communicate with one another. While humans use words and body language to communicate, neurons use electrical and chemical signals. Just like a person in a committee, one neuron usually receives and synthesizes messages from multiple other neurons before “making the decision” to send the message on to other neurons.

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 29.4**. 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**.

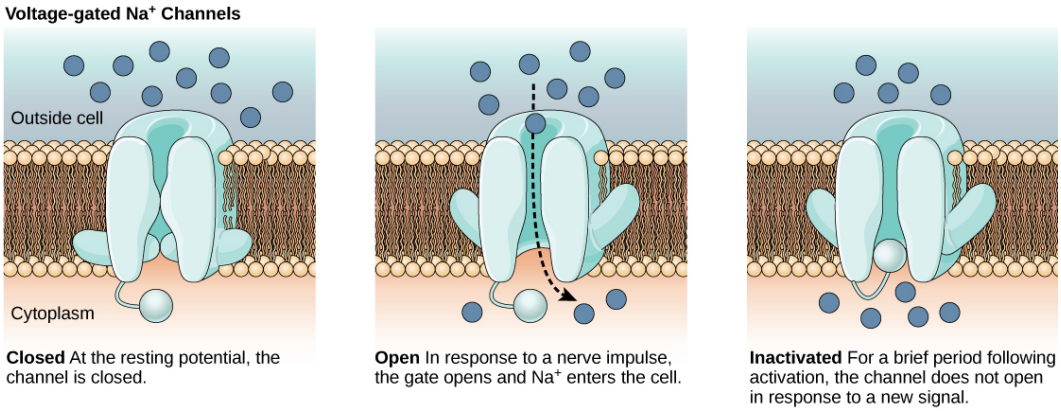


Figure 29.4 Voltage-gated ion channels open in response to changes in membrane voltage. After activation, they become inactivated for a brief period and will no longer open in response to a signal.

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 generated 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 freely cross the membrane, and because enzymes can pump ions into or out of a cell, there are different concentrations of several ions inside and outside the cell, as shown in **Table 29.1**. The difference in the number of positively charged potassium ions (K^+) inside and outside the cell dominates the resting membrane potential (**Figure 29.5**). When the membrane is at rest, K^+ ions accumulate inside the cell. 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); with more positive ions outside than inside, the inside of the cell is negatively charged (-70mV) compared to the extracellular space. The negative charge within the cell is created by the plasma 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 brings 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 chloride ions (Cl^-) tend to accumulate outside of the cell because they are repelled by negatively-charged proteins within the cytoplasm.

Ion Concentration Inside and Outside Neurons

Ion	Extracellular concentration (mM)	Intracellular concentration (mM)	Ratio outside/ inside
Na^+	145	12	12
K+	4	155	0.026
Cl^-	120	4	30
Organic anions (A−)	—	100	

Table 29.1 The resting membrane potential is a result of different concentrations inside and outside the cell.

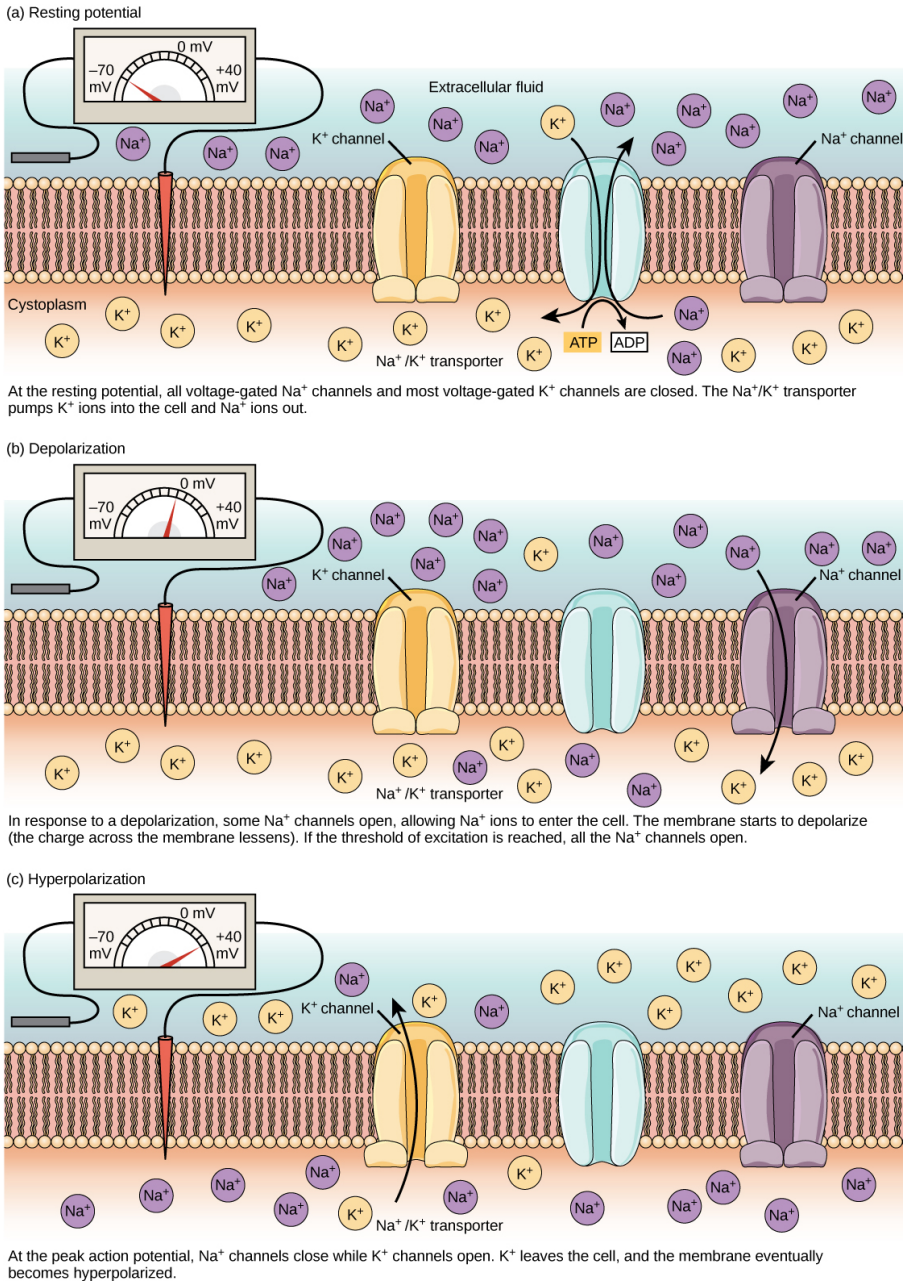


Figure 29.5 The (a) resting membrane potential is a result of different concentrations of Na^+ and K^+ ions inside and outside the cell. A nerve impulse causes Na^+ to rapidly enter the cell, resulting in (b) depolarization. At the peak action potential, K^+ channels open and the cell becomes (c) hyperpolarized.

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 dependent on chemicals we call neurotransmitters, which move between nerve cells and their targets. Transmission of a signal within a neuron (from dendrite to axon terminal) is initiated 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 29.5**

and Figure 29.6). 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.

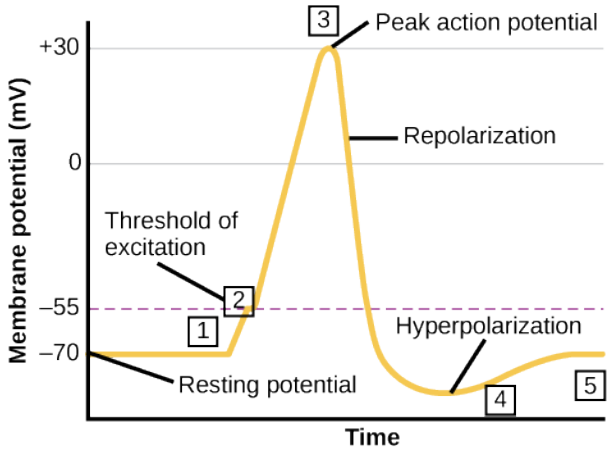


Figure 29.6 The formation of an action potential can be divided into five steps: (1) A stimulus from a sensory cell or another neuron causes the target cell to depolarize toward the threshold potential. (2) If the threshold of excitation is reached, all Na^+ channels open and the membrane depolarizes. (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.

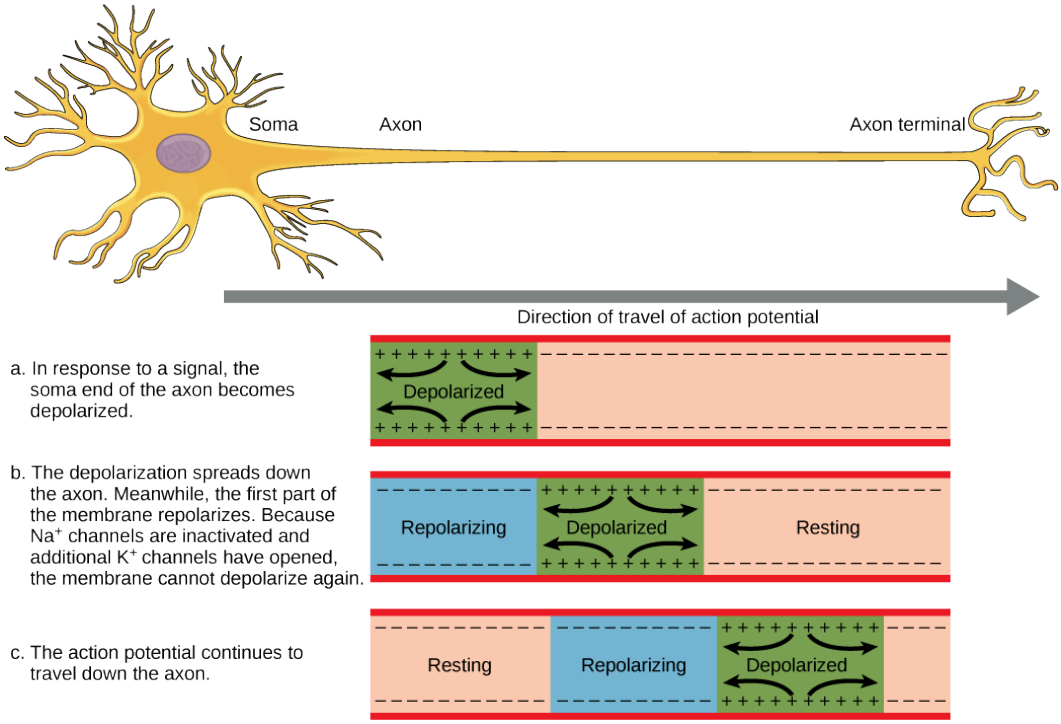


Figure 29.7 The action potential is conducted down the axon as the axon membrane depolarizes, then repolarizes.

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 Ca^{2+} 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 29.8**, which is an image from a scanning electron microscope.

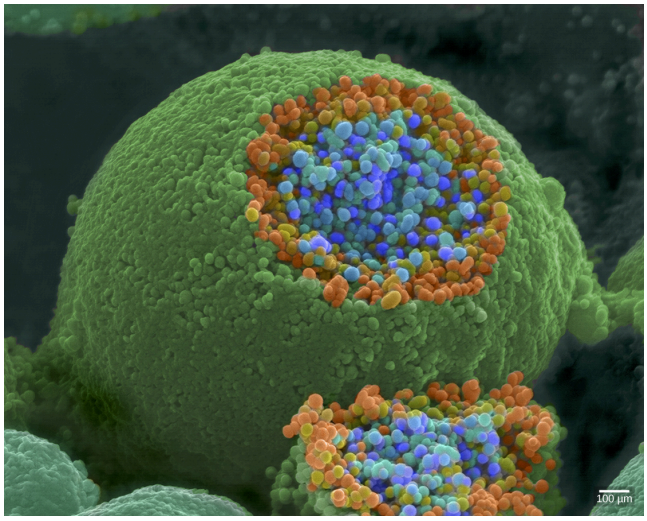


Figure 29.8 This pseudocolored image taken with a scanning electron microscope shows an axon terminal that was broken open to reveal synaptic vesicles (blue and orange) inside the neuron. (credit: modification of work by Tina Carvalho, NIH-NIGMS; scale-bar data from Matt Russell)

Fusion of a vesicle with the presynaptic neuronal plasma membrane causes neurotransmitter to be released into the **synaptic cleft**, the extracellular space between the presynaptic and postsynaptic cells, as illustrated in **Figure 29.9**. The neurotransmitter diffuses across the synaptic cleft and binds to receptor proteins on the postsynaptic cell's plasma membrane.

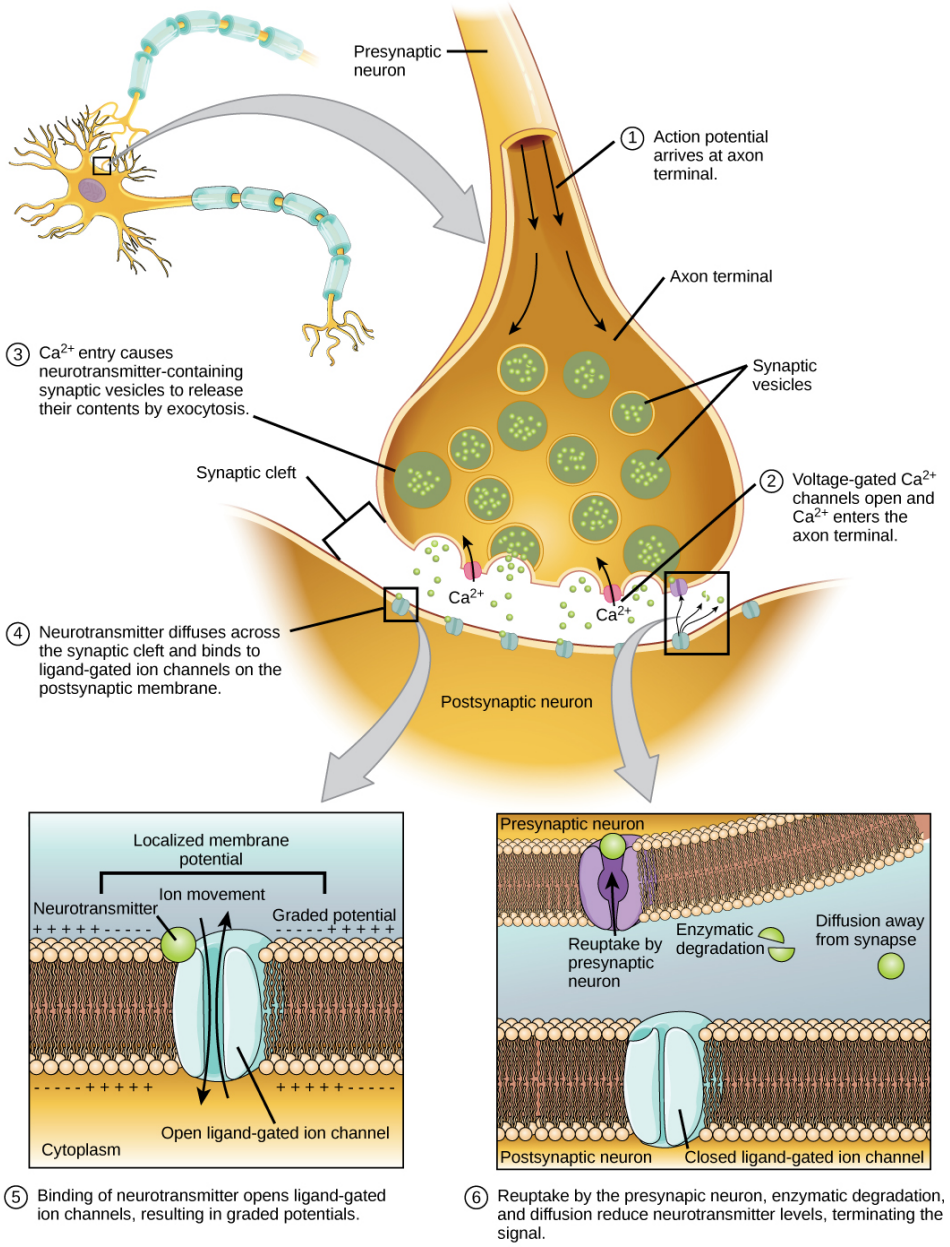


Figure 29.9 Communication at chemical synapses requires release of neurotransmitters. When the presynaptic membrane is depolarized, voltage-gated Ca^{2+} channels open and allow Ca^{2+} to enter the cell. The calcium entry causes synaptic vesicles to fuse with the membrane and release neurotransmitter 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 on the postsynaptic membrane to open. Unlike the sodium channels which respond to a change in the membrane voltage (i.e., voltage-gated channels), these ion channels are classified as **ligand-gated**, since they open the gates in response to binding of the ligand (neurotransmitter). Neurotransmitters can either have excitatory or inhibitory effects on the postsynaptic membrane, as detailed in **Table 29.1**. 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.

Once neurotransmission has occurred, the neurotransmitter must be removed from the synaptic cleft so 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 re-uptake) by the presynaptic neuron. Several drugs act at this step of neurotransmission. For example, cocaine acts to inhibit re-uptake of neurotransmitters, which acts to prolong the excitatory stimulus initiated by those neurotransmitters. Many of the well-known antidepressant drugs work in the same way.

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

Table 29.2

Electrical Synapses

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 (e.g. ATP, or signaling ions like Ca^{++}), 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.

Central Nervous System

The central nervous system (CNS) is made up of the brain, a part of which is shown in **Figure 29.10** 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 for 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.

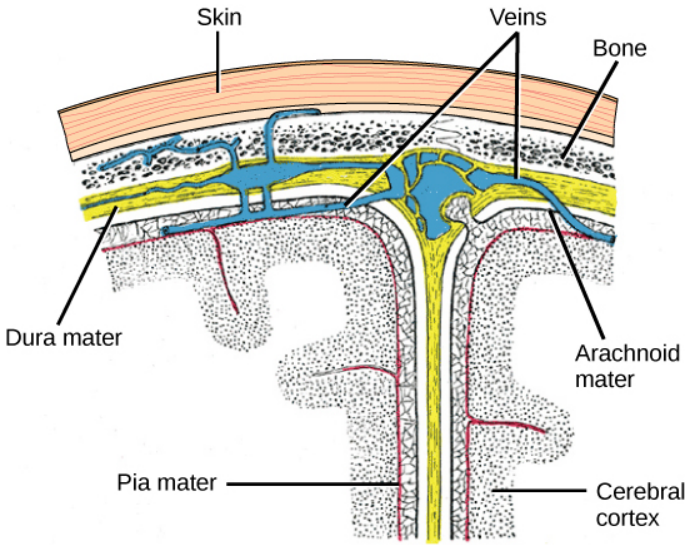


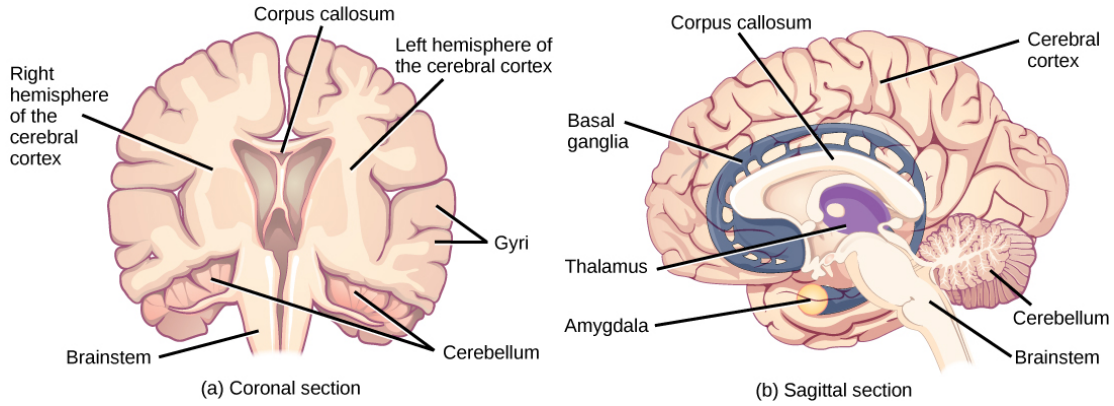
Figure 29.10 The cerebral cortex is covered by three layers of meninges: the dura, arachnoid, and pia maters. (credit: modification of work by Gray's Anatomy)

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 29.12b**, a coronal section cuts the brain front to back, as shown in **Figure 29.11a**, and a horizontal section cuts the brain top to bottom.

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.



(a) (b)

Figure 29.11 These illustrations show the (a) coronal and (b) sagittal sections of the human brain.

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, and others, have given great insight into the functions of different brain regions but have also showed 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 29.12** illustrates these four lobes of the human cerebral cortex.

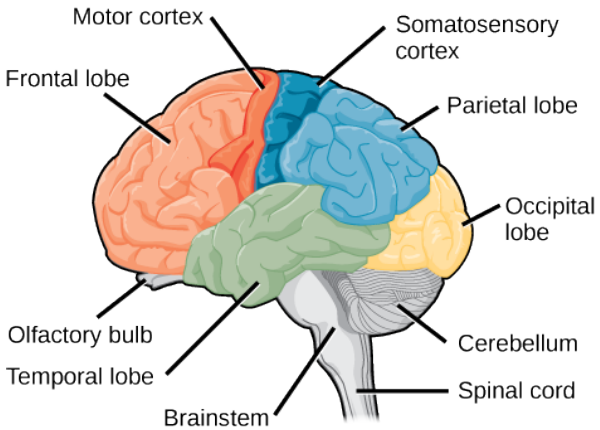


Figure 29.12 The human cerebral cortex includes the frontal, parietal, temporal, and occipital lobes.

The frontal lobe is located at the front of the brain, over the eyes. This lobe contains the olfactory bulb, which processes odors. 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 29.13**. For example, the neurons that control movement of the fingers are next to the neurons that control 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.

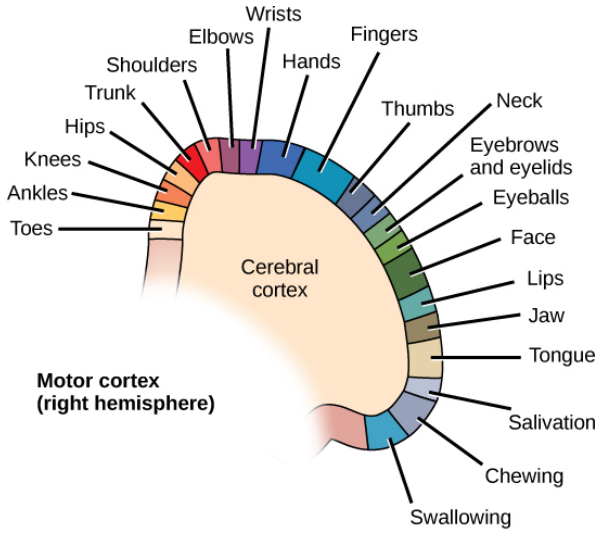


Figure 29.13 Different parts of the motor cortex control different muscle groups. Muscle groups that are neighbors in the body are generally controlled by neighboring regions of the motor cortex as well. For example, the neurons that control finger movement are near the neurons that control hand movement.

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 **m47519 (http://legacy.cnx.org/content/m47519/1.9/#fig-ch35_03_06)** . 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).

Hypothalamus

One small but critically important part of the brain is the hypothalamus, shown in **m47519 (http://legacy.cnx.org/content/m47519/1.9/#fig-ch35_03_06)** . 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 many 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.

Cerebellum

The cerebellum (Latin for "little brain"), shown in **Figure 29.12**, 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 29.12**, 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 29.12**. 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 29.14**. 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.

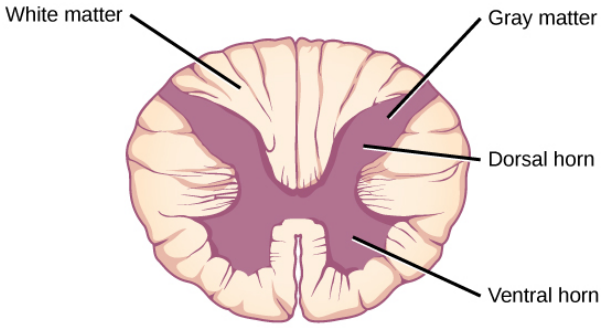


Figure 29.14 A cross-section of the spinal cord shows gray matter (containing cell bodies and interneurons) and white matter (containing axons).

Peripheral Nervous System

The peripheral nervous system (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 consists of the **sensory division** (afferent), which consists of sensory neurons, and the **motor division** (efferent), which consists of motor neurons. The sensory division conveys information to the CNS to be processed, and the motor division conveys the response of the CNA to muscles, glands and organs. The motor component of the PNS is even more complex and can be divided into the autonomic division and the somatic division. The autonomic motor division, as the name implies, is not controlled or initiated by the conscious thought of an individual. The somatic motor division is consciously controlled by the individual and usually affects skeletal muscles.

Autonomic Motor Division

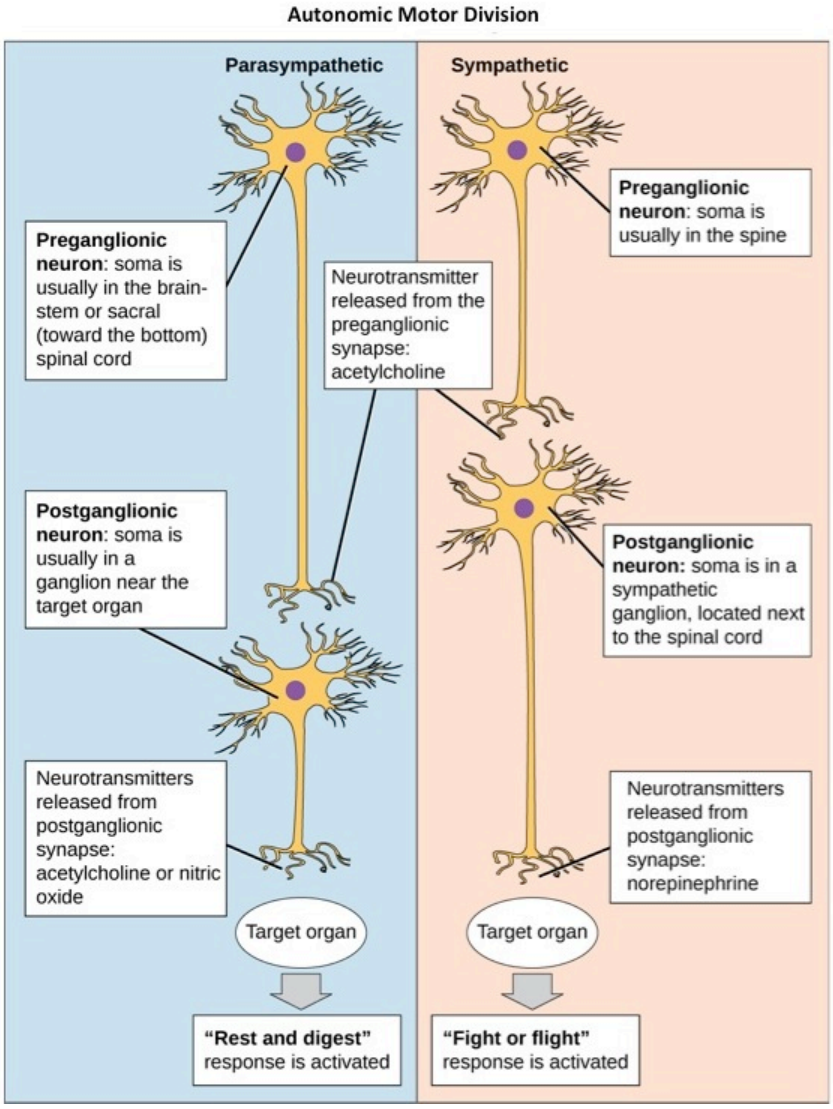


Figure 29.15 In the autonomic motor division of the peripheral nervous system, a preganglionic neuron of the CNS synapses with a postganglionic neuron of the PNS. The postganglionic neuron, in turn, acts on a target organ. Autonomic responses are mediated by the sympathetic and the parasympathetic systems, which are antagonistic to one another. The sympathetic system activates the "fight or flight" response, while the parasympathetic system activates the "rest and digest" response.

The autonomic motor division 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 motor division 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 29.15**. There are two divisions of the autonomic motor division that often have opposing effects: the sympathetic division and the parasympathetic division.

Sympathetic Division

The **sympathetic division** 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 division 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.

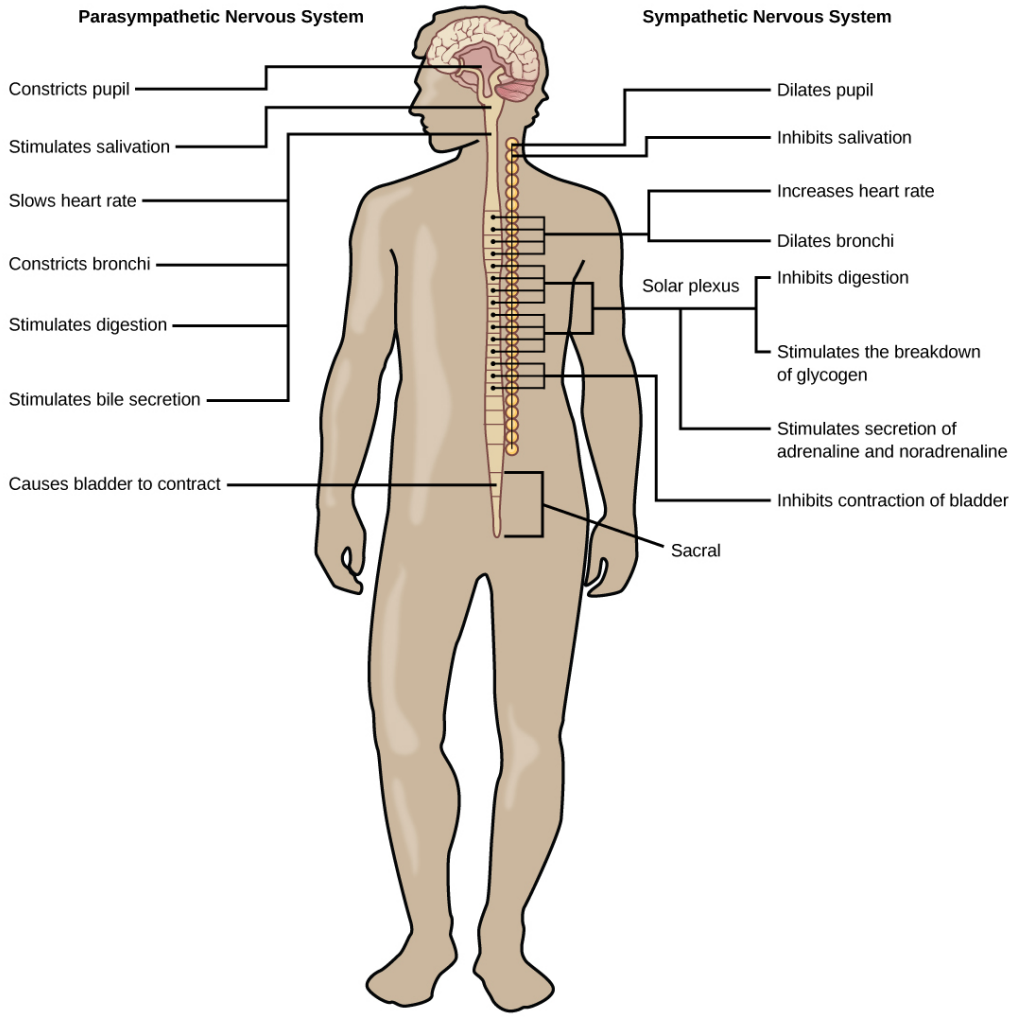


Figure 29.16 The sympathetic and parasympathetic divisions often have opposing effects on target organs.

Most preganglionic neurons in the sympathetic division originate in the spinal cord, as illustrated in **Figure 29.16**. The axons of these neurons release **acetylcholine** on postganglionic neurons within 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 division 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 blood stream. 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 helps an organism avoid danger.

Parasympathetic Division

While the sympathetic division is activated in stressful situations, the **parasympathetic division** 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 division 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 29.16**. 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 division resets organ function after the sympathetic division 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.

Somatic Division

The somatic division of the peripheral nervous system is made up of cranial and spinal nerves the contain motor neurons. Motor neurons transmit messages about desired movement from the CNS to the muscles to make them contract. Without its somatic division, 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 division, which has two synapses between the CNS and the target organ, motor neurons of the somatic division have only one synapse between the CNS and muscle or organ. Acetylcholine is the main neurotransmitter released at these synapses.

Cranial and Spinal Nerves

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 are detailed in **Figure 29.17**. 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).

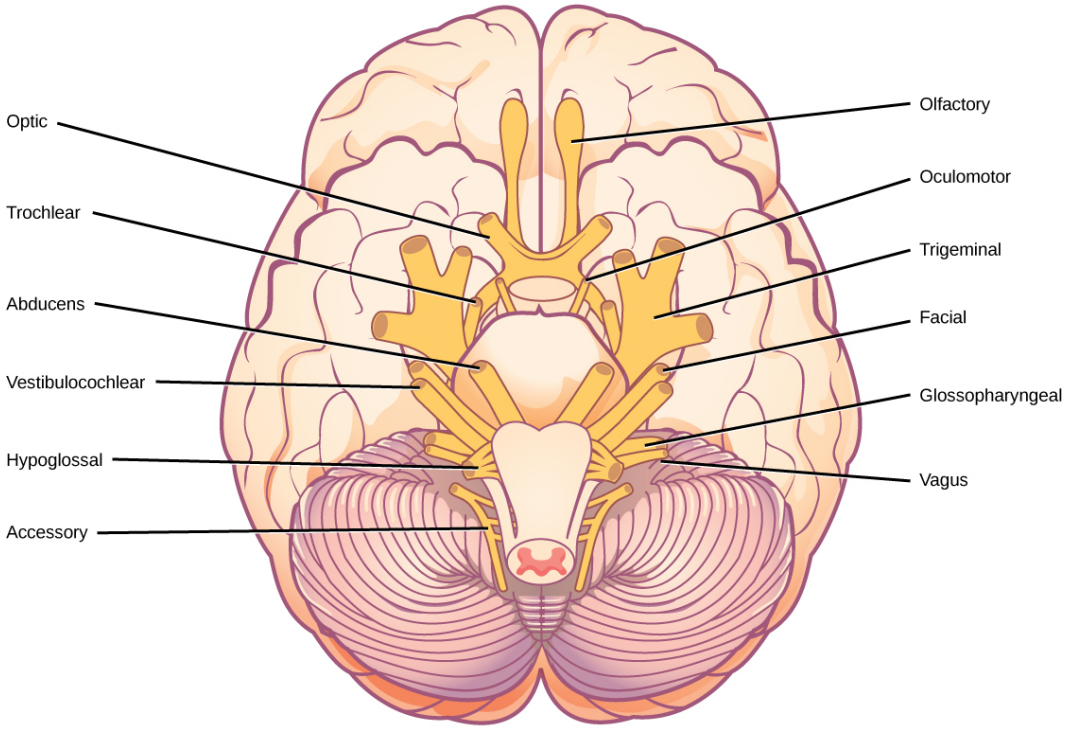


Figure 29.17 The human brain contains 12 cranial nerves that receive sensory input and control motor output for the head and neck.

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 29.18**. Each sensory neuron has one projection—with a sensory receptor ending in 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.

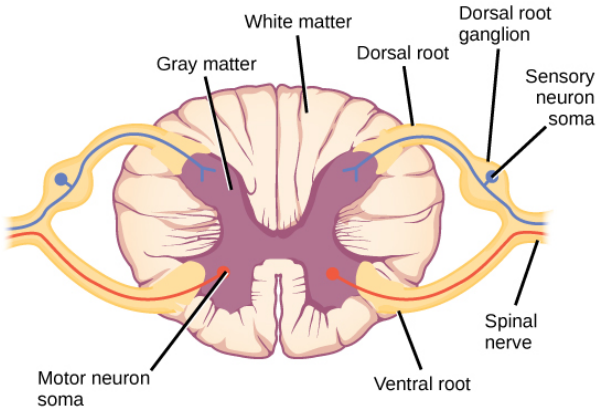


Figure 29.18 Spinal nerves contain both sensory and motor axons. The somas of sensory neurons are located in dorsal root ganglia. The somas of motor neurons are found in the ventral portion of the gray matter of the spinal cord.

29.2 | Sensory Systems

Introduction

> “ The act of smelling something, anything, is remarkably like the act of thinking. Immediately at the moment of perception, you can feel the mind going to work, sending the odor around from place to place, setting off complex repertories through the brain, polling one center after another for signs of recognition, for old memories and old connection.”

Lewis Thomas, *On Smell*, 1985

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**. We are going to focus on the five special senses of humans.

There are three types of stimuli that are detected by the human sensory systems. The first is **chemical stimulus**, where molecules stimulate a sensory neuron. Chemical stimuli are detected by the olfactory system, when molecules in the air bind to sensory cells in the nasal epithelia, and in the gustation (taste) system, when molecules in your food stimulate your taste buds. The second is **electromagnetic radiation**; light interacts with molecules in the sensory cells (rods and cones) of your retina, and those sensory cells send a signal to your brain. The third is **mechanical stimulation**, where the sensory cells are activated by movement or touch. Mechanical stimuli are detected by the cells in the inner ear that help you detect balance and body position, and by other cells in your inner ear detect sound (the sense of hearing). Additionally, mechanical stimuli are involved in other somatosensory systems, such as pressure, pain, or vibration, in proprioception (position of legs, arms and other body parts), and in kinesthesia (detection of motion of those same body parts.

Sensory Perception

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 a produces a change in electrical potential in response to the stimulus. This is called the receptor potential. How is sensory input, such as pressure on the skin, changed to a receptor potential? In one example, a type of receptor called a **mechanoreceptor** (as shown in **Figure 29.19**) 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 all cases the appropriate stimulus will cause a change in the membrane potential of the sensory cell; the exact mechanism for changing the membrane potential will be different for different sensory calls.

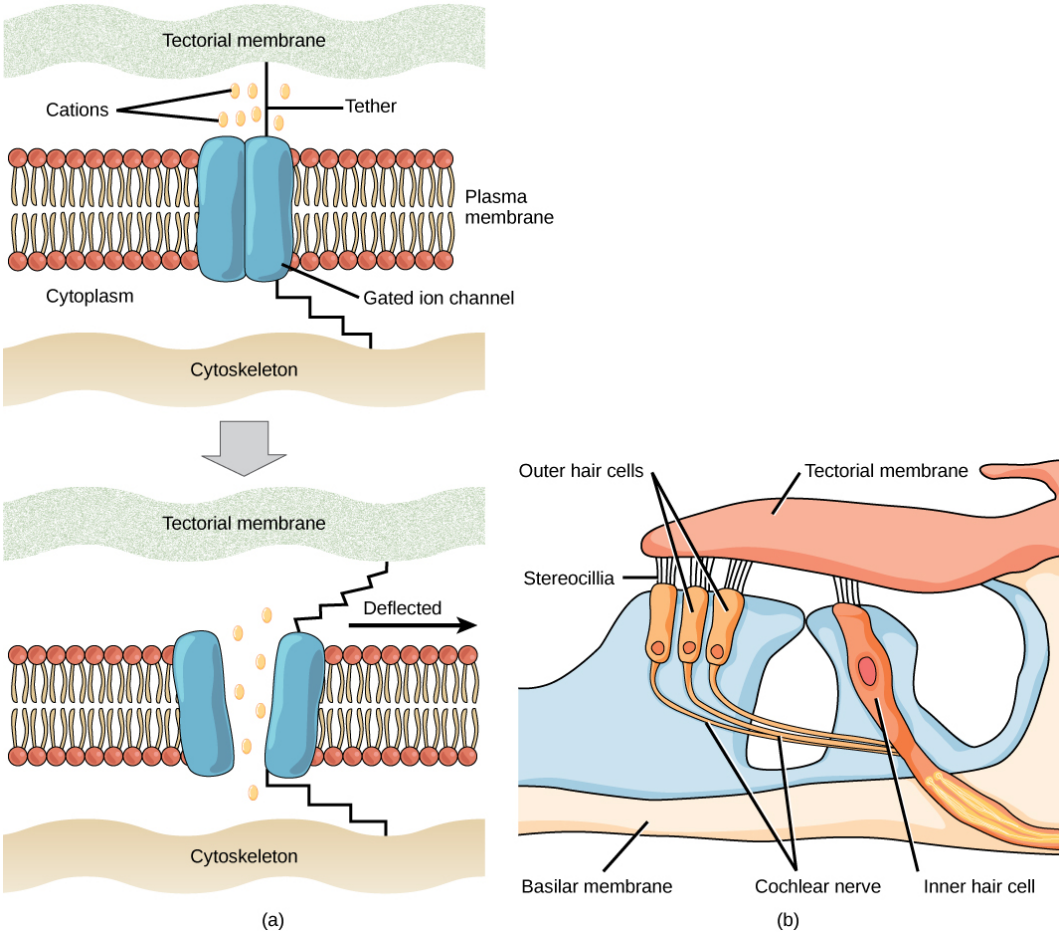


Figure 29.19 (a) Mechanosensitive ion channels are gated ion channels that respond to mechanical deformation of the plasma membrane. A mechanosensitive channel is connected to the plasma membrane and the cytoskeleton by hair-like tethers. When pressure causes the extracellular matrix to move, the channel 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.

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 29.20**) 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. Interestingly, studies have shown that the allele that results in red hair in humans homozygous for that allele (known as MC1R) is a member of the family of sensory receptors that detect pain. And redheads are more sensitive to pain, and require about 20%

more anesthetic during surgery or dental work. So be nice to your red-headed friends.

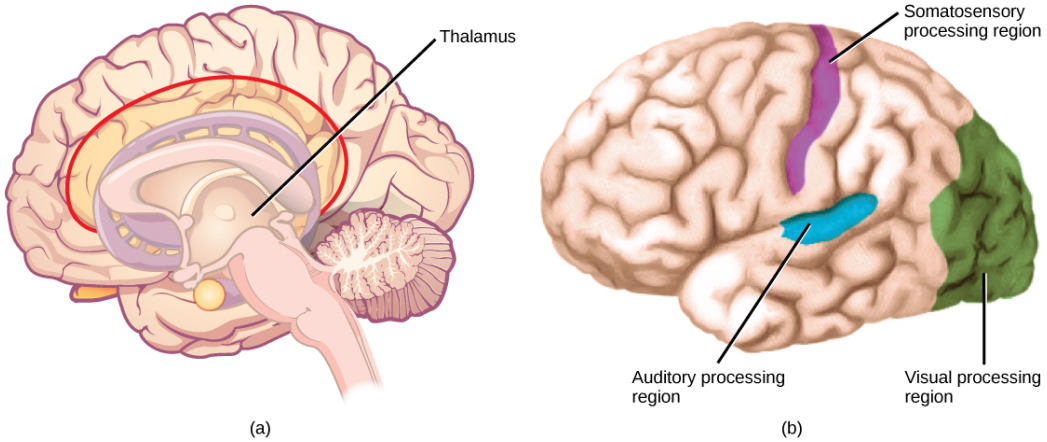


Figure 29.20 In humans, with the exception of olfaction, all sensory signals are routed from the (a) thalamus to (b) final processing regions in the cortex of the brain. (credit b: modification of work by Polina Tishina)

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 volatile chemicals 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

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 29.21**).

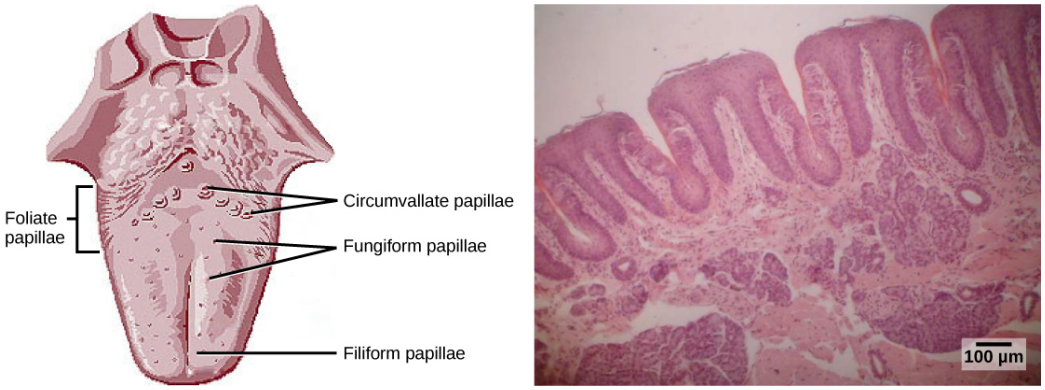


Figure 29.21 (a) Foliate, circumvallate, and fungiform papillae are located on different regions of the tongue. (b) Foliate papillae are prominent protrusions on this light micrograph. (credit a: modification of work by NCI; scale-bar data from Matt Russell)

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 29.22**). 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.

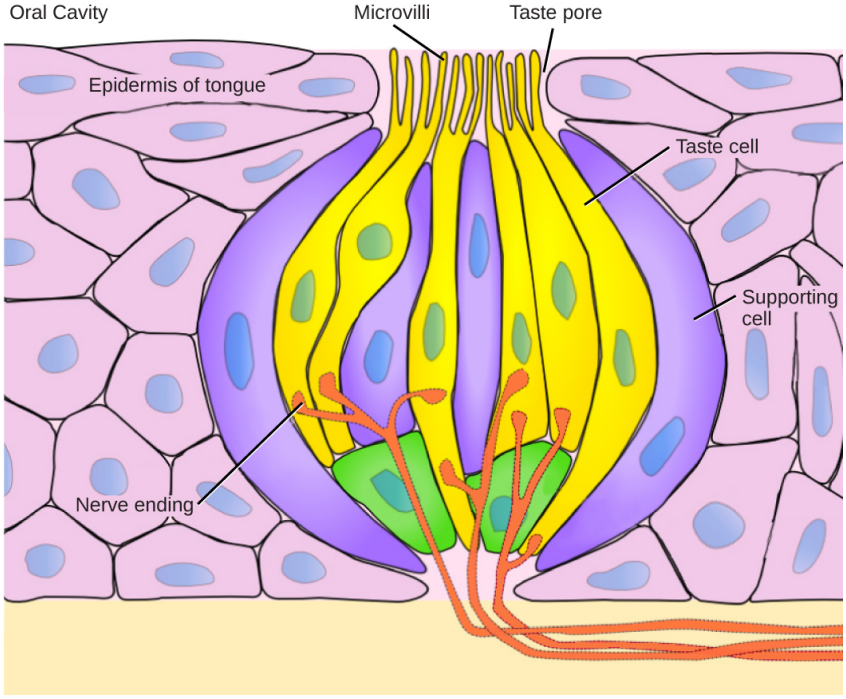


Figure 29.22 Pores in the tongue allow tastants to enter taste pores in the tongue. (credit: modification of work by Vincenzo Rizzo)

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). 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.

Hearing and Equilibrium

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.

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 29.23**). 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 have been adapted for use in the vertebrate ear over evolutionary time.

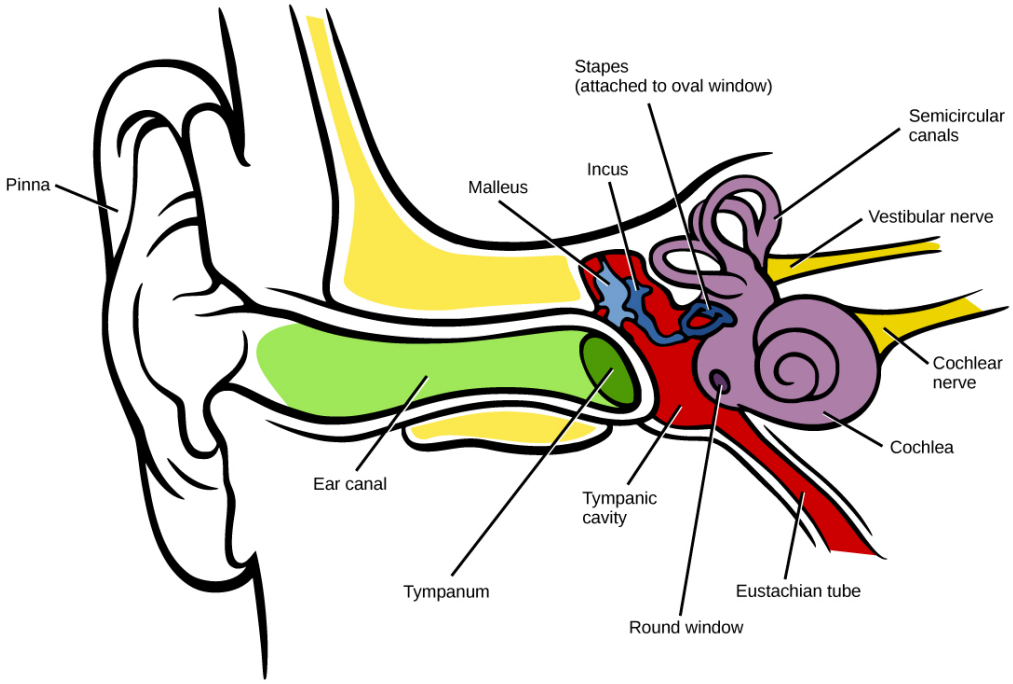


Figure 29.23 Sound travels through the outer ear to the middle ear, which is bounded on its exterior by the tympanic membrane. The middle ear contains three bones called ossicles that transfer the sound wave to the oval window, the exterior boundary of the inner ear. The organ of Corti, which is the organ of sound transduction, lies inside the cochlea. (credit: modification of work by Lars Chittka, Axel Brockmann)

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 29.24**. 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.

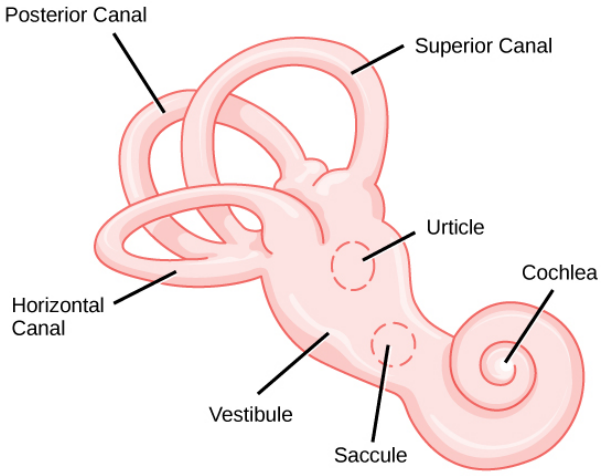


Figure 29.24 The structure of the vestibular labyrinth is shown. (credit: modification of work by NIH)

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 29.24**. 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.

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 29.25**). 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).

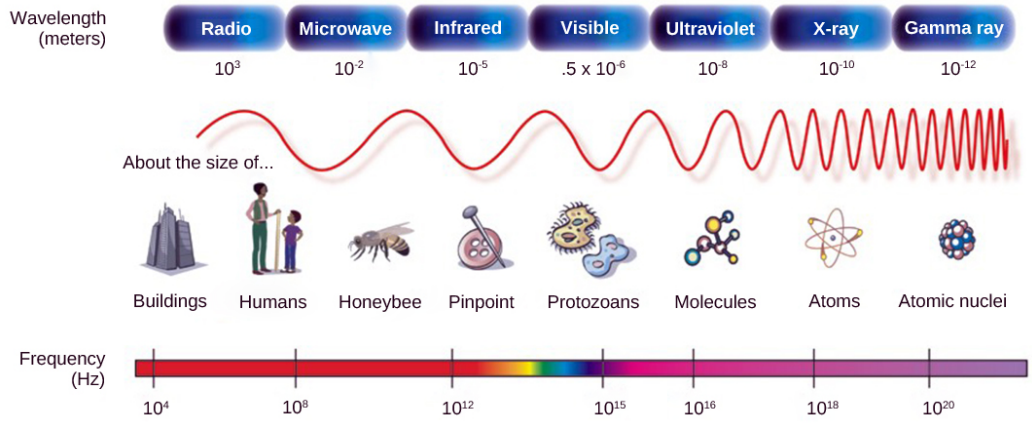


Figure 29.25 In the electromagnetic spectrum, visible light lies between 380 nm and 740 nm. (credit: modification of work by NASA)

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 (shown in **Figure 29.26**) 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 29.26b**). 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.

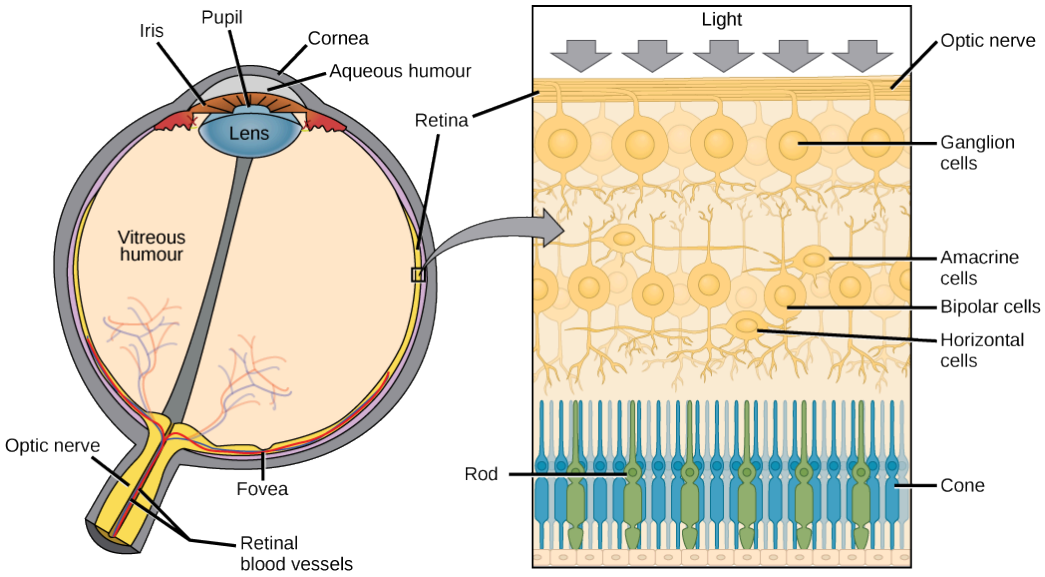


Figure 29.26 (a) The human eye is shown in cross section. (b) A blowup shows the layers of the retina.

The main function of the lens is to focus light on the retina and fovea centralis. The lens is dynamic, focusing and re-focusing light as the eye rests on near and far objects in the visual field. The lens is operated by muscles that stretch it flat or allow it to thicken, changing the focal length of light coming through it to focus it sharply on the retina. With age comes the loss of the flexibility of the lens, and a form of farsightedness called presbyopia results. Presbyopia occurs because the image focuses behind the retina. Presbyopia is a deficit similar to a different type of farsightedness called hyperopia caused by an eyeball that is too short. For both defects, images in the distance are clear but images nearby are blurry. Myopia (nearsightedness) occurs when an eyeball is elongated and the image focus falls in front of the retina. In this case, images in the distance are blurry but images nearby are clear.

There are two types of photoreceptors in the retina: **rods** and **cones**, named for their general appearance as illustrated in **Figure 29.27**. 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.

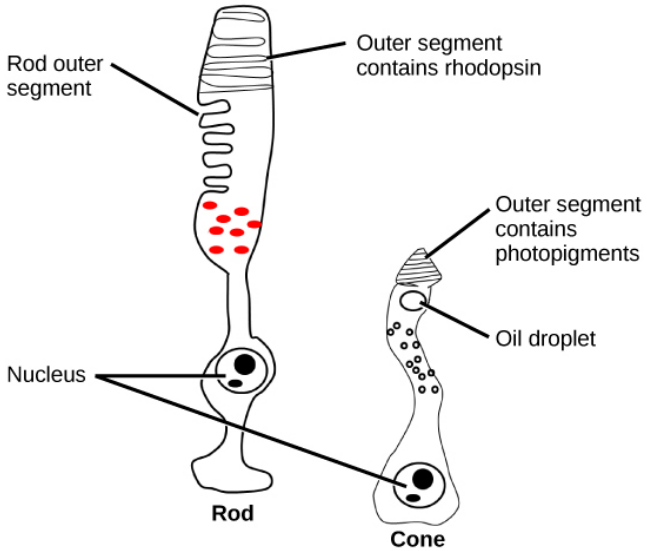


Figure 29.27 Rods and cones are photoreceptors in the retina. Rods respond in low light and can detect only shades of gray. Cones respond in intense light and are responsible for color vision. (credit: modification of work by Piotr Sliwa)

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.

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 a change in the membrane potential of the rod or cone cell.

Trichromatic Coding

There are three types of cones (with different photopsins), and they differ in the wavelength to which they are most responsive, as shown in **Figure 29.28**. Some cones are maximally responsive to short light waves of 420 nm, so they are called S cones (“S” for “short”); others respond maximally to waves of 530 nm (M cones, for “medium”); a third group responds maximally to light of longer wavelengths, at 560 nm (L, or “long” cones). With only one type of cone, color vision would not be possible, and a two-cone (dichromatic) system has limitations. Primates use a three-cone (trichromatic) system, resulting in full color vision.

The color we perceive is a result of the ratio of activity of our three types of cones. The colors of the visual spectrum, running from long-wavelength light to short, are red (700 nm), orange (600 nm), yellow (565 nm), green (497 nm), blue (470 nm), indigo (450 nm), and violet (425 nm). Humans have very sensitive perception of color and can distinguish about 500 levels of brightness, 200 different hues, and 20 steps of saturation, or about 2 million distinct colors.

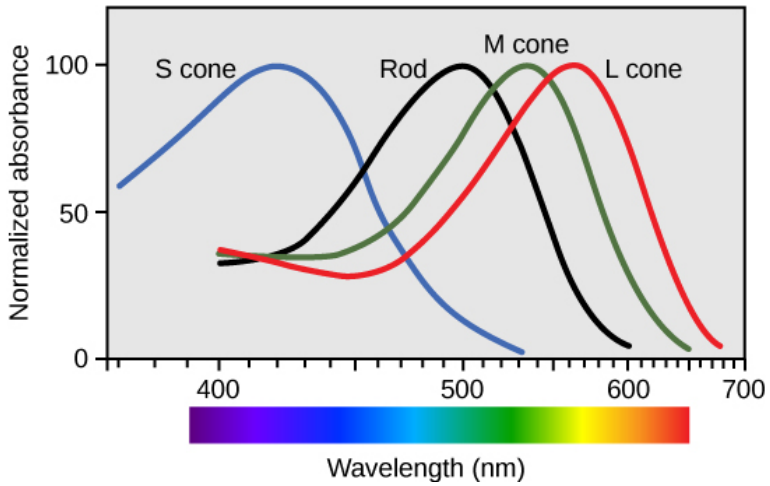


Figure 29.28 Human rod cells and the different types of cone cells each have an optimal wavelength. However, there is considerable overlap in the wavelengths of light detected.

29.3 | Musculoskeletal System

Introduction

> “Of all the constituents of the human body, bone is the hardest, the driest, the earthiest and the coldest; and, excepting only the teeth, it is devoid of sensation. God, the great Creator of all things, formed its substance to this specification with good reason, intending it to be like a foundation for the whole body; for it the fabric of the human

> body bones perform the same function as do walls and beams in houses, poles in tents, and keels and ribs in boats.”

Andreas Vesalius, Flemish anatomist, in *De Humanis Corporis Fabrica*, 1543

As Vesalius recognized long ago, the muscular and skeletal systems provide support to the body and allow for a wide range of movement. The bones of the skeletal system 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. Injury or disease affecting the musculoskeletal system can be very debilitating. In humans, the most common musculoskeletal diseases worldwide are caused by malnutrition. Ailments that affect the joints are also widespread, such as arthritis, which can make movement difficult and—in advanced cases—completely impair mobility.

Types of Skeletal Systems

A skeletal system is necessary to support the body, protect internal organs, and allow for the movement of an organism. There are three different skeleton designs that fulfill these functions: hydrostatic skeleton, exoskeleton, and endoskeleton.

Hydrostatic Skeleton

A **hydrostatic skeleton** is a skeleton formed by a fluid-filled compartment within the body, called the coelom. The organs of the coelom are supported by the aqueous fluid, which also resists external compression. This compartment is under hydrostatic pressure because of the fluid and supports the other organs of the organism. This type of skeletal system is found in soft-bodied animals such as sea anemones, earthworms, Cnidaria, and other invertebrates (**Figure 29.29**).

Figure 29.29 The skeleton of the red-knobbed sea star (*Protoreaster linckii*) is an example of a hydrostatic skeleton. (credit: “Amada44”/Wikimedia Commons)

Movement in a hydrostatic skeleton is provided by muscles that surround the coelom. The muscles in a hydrostatic skeleton contract to change the shape of the coelom; the pressure of the fluid in the coelom produces movement. For example, earthworms move by waves of muscular contractions of the skeletal muscle of the body wall hydrostatic skeleton, called peristalsis, which alternately shorten and lengthen the body. Lengthening the body extends the anterior end of the organism. Most organisms have a mechanism to fix themselves in the substrate. Shortening the muscles then draws the posterior portion of the body forward. Although a hydrostatic skeleton is well-suited to invertebrate organisms such as earthworms and some aquatic organisms, it is not an efficient skeleton for terrestrial animals.

Exoskeleton

An **exoskeleton** is a chitinous external skeleton that consists of a hard encasement on the surface of an organism. For example, the shells of crabs and insects are exoskeletons (**Figure 29.30**). This skeleton type provides defense against predators, supports the body, and allows for movement through the contraction of attached muscles. As with endoskeletons (see below), muscles must cross a joint inside the exoskeleton. Shortening of the muscle thus changes the relationship of the two segments of the exoskeleton. Arthropods such as crabs and lobsters have exoskeletons that consist of 30–50 percent chitin, a polysaccharide derivative of glucose that is a strong but flexible material. Chitin is secreted by the epidermal cells. The exoskeleton is further strengthened by the addition of calcium carbonate in organisms such as the lobster. Because the exoskeleton is acellular, arthropods must periodically shed their exoskeletons as they grow, because the exoskeleton does not grow as the organism grows.

Figure 29.30 Muscles attached to the exoskeleton of the Halloween crab (*Gecarcinus quadratus*) allow it to move.

Endoskeleton

An **endoskeleton** is a skeleton that consists of hard, mineralized structures located within the soft tissue of organisms. The bones of vertebrates are composed of tissues and mineralized tissues. Endoskeletons provide support for the body, protect internal organs, and allow for movement through contraction of muscles attached to the skeleton.

Figure 29.31 The skeletons of humans and horses are examples of endoskeletons. (credit: Ross Murphy)

As an example, the human skeleton is an endoskeleton that consists of 206 bones in the adult. It 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 the shoulders, limb bones, the pectoral girdle, and the pelvic girdle).

Main divisions of the vertebrate skeleton

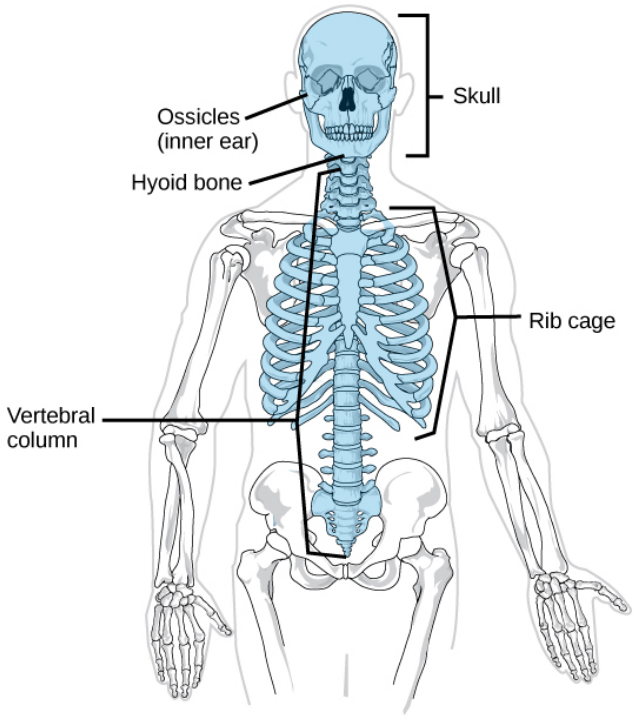


Figure 29.32 The axial skeleton of humans consists of the bones of the skull, ossicles of the middle ear, hyoid bone, vertebral column, and rib cage. (credit: modification of work by Mariana Ruiz Villareal)

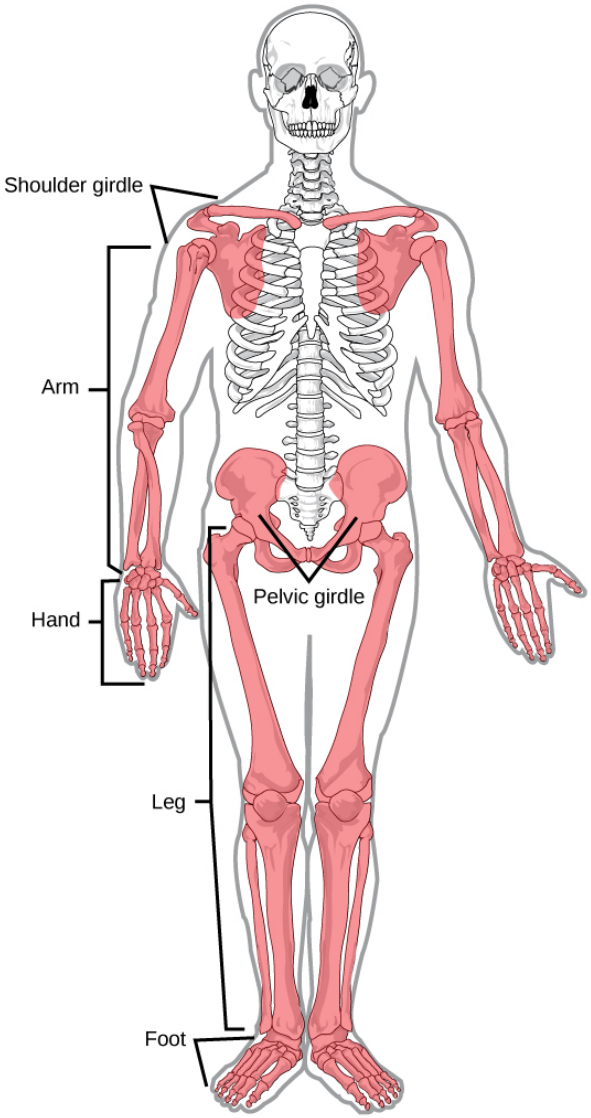


Figure 29.33 The human appendicular skeleton is composed of the bones of the pectoral limbs (arm, forearm, hand), the pelvic limbs (thigh, leg, foot), the pectoral girdle, and the pelvic girdle. (credit: modification of work by Mariana Ruiz Villareal)

Evolution of Body Design for Locomotion on Land

The transition of vertebrates from fish ancestors to land-dwelling animals required a number of changes in body design, since movement on land poses challenges that are different from those posed by movement in water. Water provides a certain amount of lift, which is missing on land, so muscles are needed to provide that lift on land. It also provides a medium to push against, and many fish use lateral undulations to push against the water and generate forward movement. Air does not provide the same resistance, and so lateral undulations on land would not produce much movement forward.

As certain fish began to move onto land, they retained their lateral undulation form of locomotion. However, instead of pushing against water, their fins became points of contact with the ground, and the lateral undulations became rotations about those points of contact. The lack of bouyancy provided by water led to increased ossification of the bones and strengthening of the muscles, as well as providing selective pressure that resulted in changes in arrangement of the "wrist" bones of these early tetrapods. The effect of gravity also led to changes in the axial skeleton, since rotations around the contact points with the ground caused new torsional strains on the vertebral column. A firmer, more ossified vertebral column became common in land animals, because it reduces the strain and also provides strength to support the weight of the body. In later tetrapods the vertebrae began allowing for vertical rather than lateral flexing. The vertebrae of the neck also evolved to allow movement of the head independently of the body, a range of motion not found in fish.

In early terrestrial tetrapods (**Figure 3.3**), the limbs were splayed out to the side, reflecting the position of the fins in their

fishy ancestors. This resulted in a form of motion that was similar to performing push-ups while walking, which requires large muscles to move the limbs toward the midline. This is not an efficient form of locomotion, and selective pressures soon led to a configuration where the limbs were placed underneath the body, so that each stride requires less energy to move the animal forward. The rotation around the point of contact became a motion that is more like a pendulum when the limbs are underneath the body, producing a stride that was much more efficient for movement over land. Additional changes were required in the appendicular skeleton to accommodate the new ranges of motion that were enabled by that limb placement.

Muscles

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 29.34).

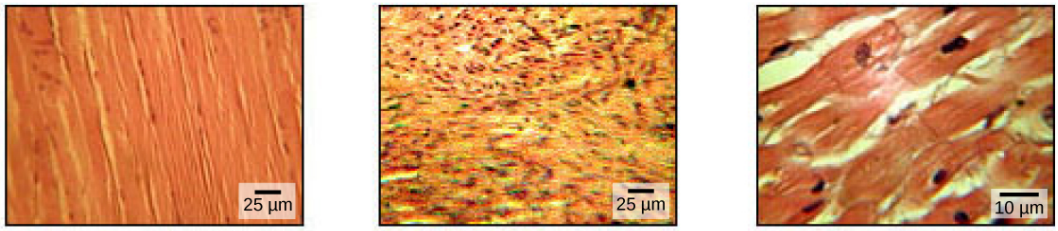


Figure 29.34 The body contains three types of muscle tissue: skeletal muscle, smooth muscle, and cardiac muscle, visualized here using light microscopy. Smooth muscle cells are short, tapered at each end, and have only one plump nucleus in each. Cardiac muscle cells are branched and striated, but short. 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 cells form by fusion of many muscle cells called myoblasts, and thus have 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 also 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. **Intercalated disks** are ion-permeable junctions between individual cardiac muscle cells which allow for synchronized contractions of the various regions of the heart.

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 29.35).

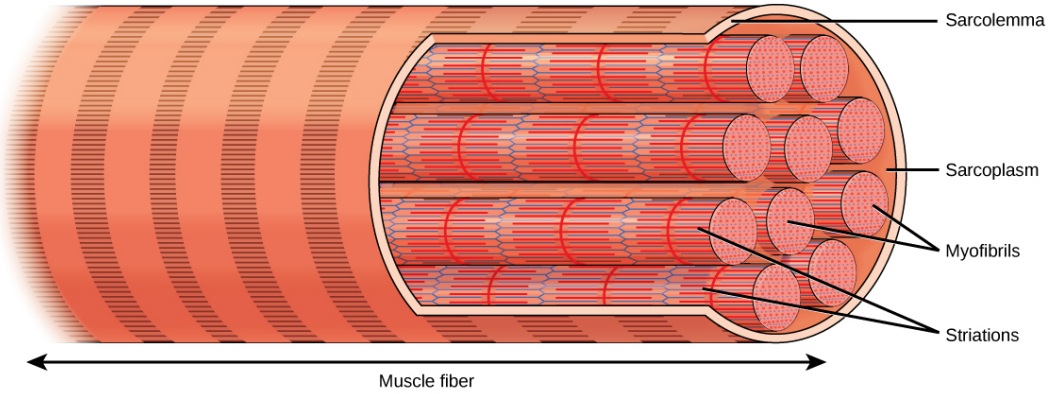


Figure 29.35 A skeletal muscle cell is surrounded by a plasma membrane called the sarcolemma with a cytoplasm called the sarcoplasm. A muscle fiber is composed of many fibrils, packaged into orderly units.

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 29.36**).

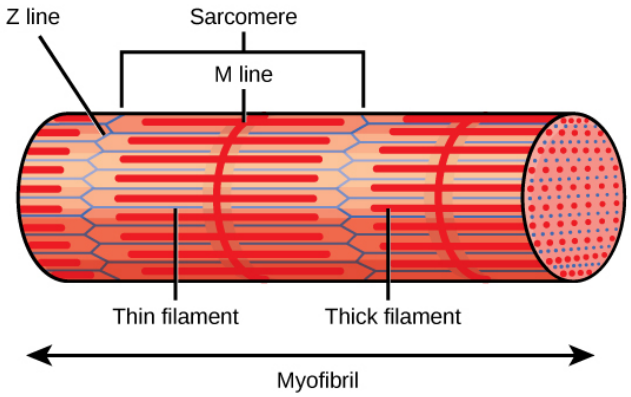


Figure 29.36 A sarcomere is the region from one Z line to the next Z line. Many sarcomeres are present in a myofibril, resulting in the striation pattern characteristic of skeletal muscle.

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 primarily 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 Ca^{2+} ions.

Sliding Filament Model of Contraction

For a muscle cell to contract, the sarcomere must shorten. However, individual 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 explain the differences observed in the lengths of 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 29.37**).

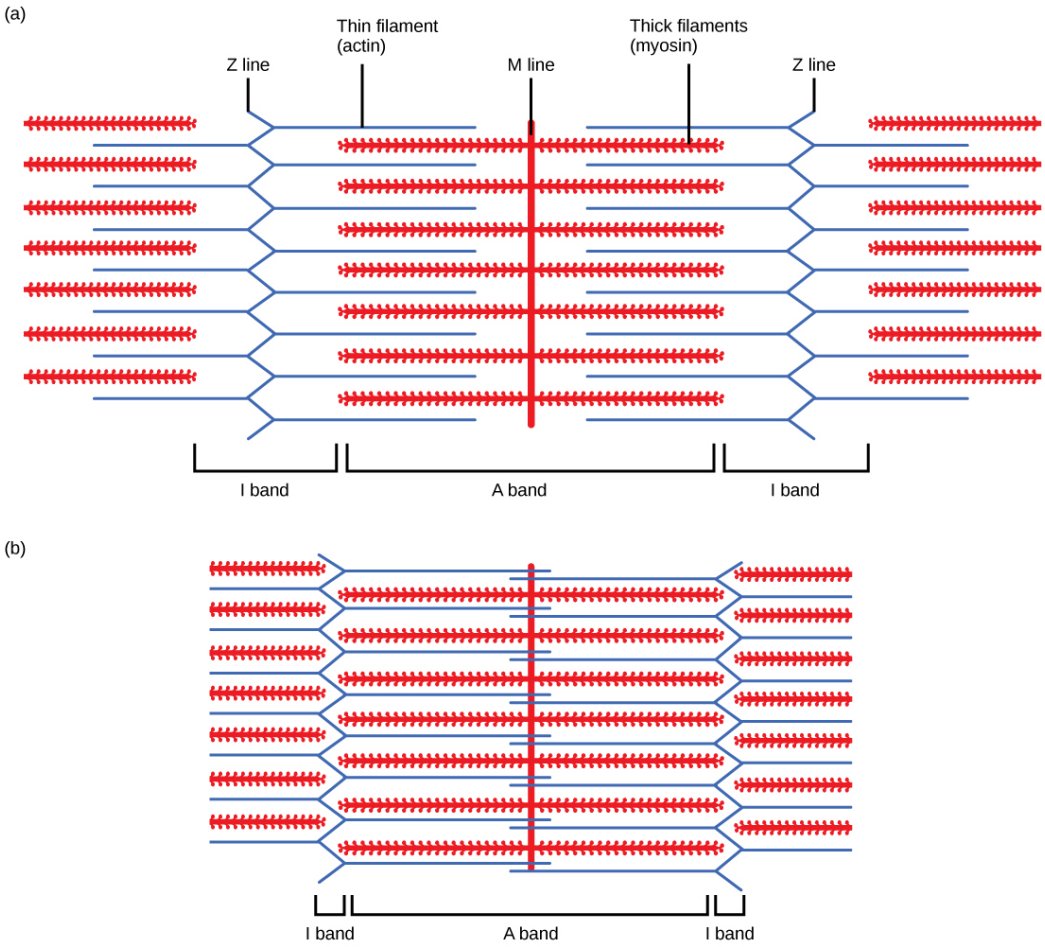


Figure 29.37 When (a) a sarcomere (b) contracts, the Z lines move closer together and the I band gets smaller. The A band stays the same width and, at full contraction, the thin filaments overlap.

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, and acts as an enzyme to convert ATP to ADP, releasing an inorganic phosphate molecule and energy. The energy can be harnessed to promote contraction via the sliding filament mechanism described above.

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, P_i. 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 P_i 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. P_i 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 29.38**).

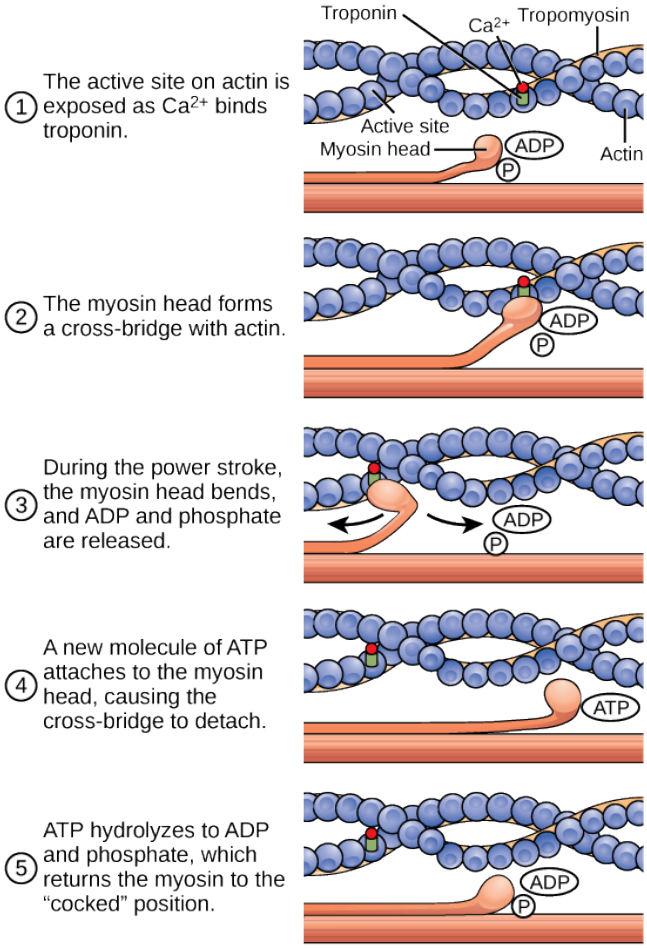


Figure 29.38 The cross-bridge muscle contraction cycle, which is triggered by Ca^{2+} binding to the actin active site, is shown. With each contraction cycle, actin moves relative to myosin, and the thick and thin filaments slide past each other.

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.

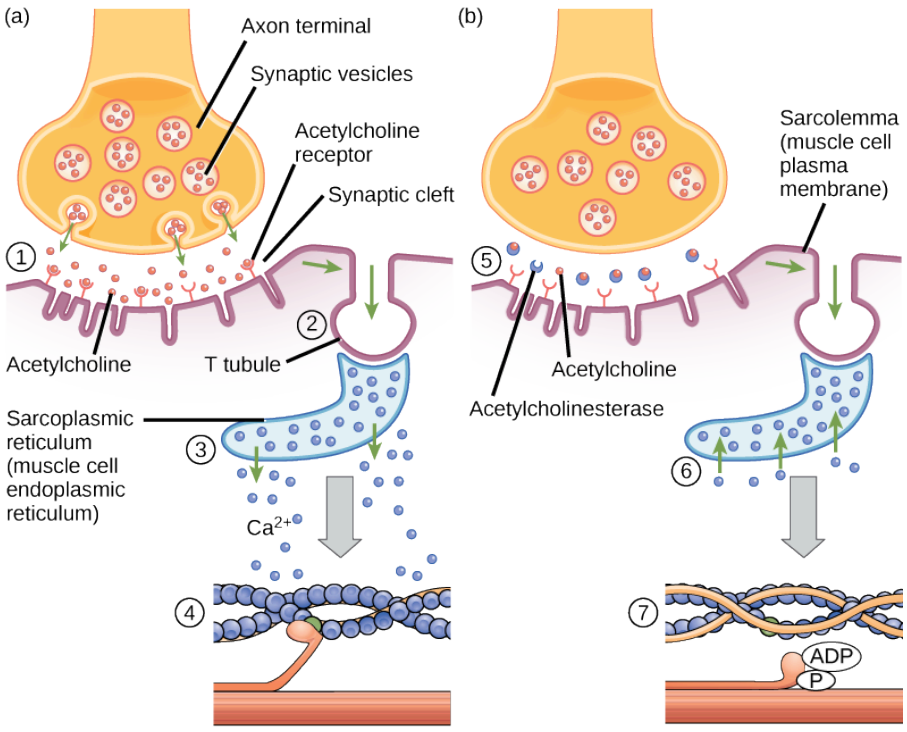
To enable a muscle contraction, tropomyosin must change conformation, uncovering the myosin-binding site on an actin molecule and allowing cross-bridge formation. This can only happen in the presence of calcium, which is kept at extremely low concentrations in the sarcoplasm. If present, calcium ions 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 Ca^{2+} 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. This junction is functionally similar to a synapse between two nerve cells, allowing a signal from the nerve cell to initiate an action potential in the muscle plasma membrane. The action potential in the muscle cell causes Ca^{++} to be released from intracellular stores; this elevated calcium concentration triggers the binding of actin and myosin, ATP hyrolysis, and all of the other steps in contraction that are outlined above.

The neurotransmitter released at the neuromuscular junction in most animals is acetylcholine. It is released from the nerve call ending and binds to receptors on the muscle cell plasma membrane (**Figure 29.39**); these receptors act as sodium channels when acetylcholine is bound to them. The influx of sodium depolarizes the muscle cell, triggering an action potential in a similar fashion to the action potential found in nerve cells. The acetylcholine is rapidly degraded in the neuromuscular junction by an enzyme called acetylcholinesterase. Various natural toxins (such as the curare used on poison arrows by South American indigenous tribes) and synthetic toxins (including nerve gases and insecticides) target the components of the neuromuscular junction, including both the receptor and the acetylcholinesterase. The deadly nerve gas known as Sarin irreversibly inhibits acetylcholinesterase. What effect would Sarin have on muscle contraction, and how does that effect lead to death?



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 29.39 Acetylcholine effects at the neuromuscular junction. The depolarization of the muscle plasma membrane releases calcium from the sarcoplasmic reticulum and initiates contraction of the muscle.

29.4 | Reflexes and Homeostasis

Introduction

> "There is no other species on Earth that does science. It is, so far, entirely a human invention, evolved by natural selection in the human cortex for one simple reason: it works. It is not perfect. It is misused. It is only a tool. But it is by far the best tool we have, self-correcting, ongoing, applicable to everything."

Carl Sagan, American astronomer, in *Cosmos*, 1980

As we come to the last section of this book, it seems appropriate to hearken back to where we started - with a reminder that science is a way of knowing. The knowledge of animal biology discussed in this module, and the mechanisms of nervous system function, homeostasis, etc. are all products of painstaking experiments and observations, in some cases dating back hundreds of years. This uniquely human endeavor gave us that knowledge. More importantly, animal biology is just a subset of the science of biology, and biology is just a subset of science. There is a lot more to explore in other areas of science, and

infinitely more for current and future humans to learn about all of them. We hope that the knowledge and thinking processes that you have used in this course will serve you well in your future education, and will help you attain success in whatever endeavor you choose for your future work.

Reflexes

When the body reacts involuntarily to an internal or external stimulus, the response is called a reflex, and the neurons that make up the simple circuit are called a **reflex arc**. This is involuntary, spontaneous, and does not involve processing of stimuli by higher centers (e.g. regions of the brain). Reflexes can be spinal or cranial, depending on the nerves and central components that are involved. For example heat and pain sensations from a hot stove causing withdrawal of the arm through a connection in the spinal cord that leads to contraction of the muscles in the arm, jerking the arm away from the hot stove.

Other examples of reflexes include

- **Withdrawal reflex**, which occurs when you step on a painful stimulus, like a tack or a sharp rock. The pain receptors (nociceptors) that are activated by the painful stimulus activate the motor neurons responsible for contraction of the leg muscles to remove your foot from the pain.
- **Stretch reflex**, which helps to maintain muscles and an optimal length. Receptors (called spindle receptors) within muscles are activated when the muscle is stretched, causing direct contraction of the muscle.
- **Corneal reflex**, or eye-blink reflex. When the cornea is stimulated, whether by touch or a bright light, blinking of the eyelids is initiated. Obviously this is to keep the cornea safe from abrasion by dust or flying insects, etc, and to protect the lens and retina from over-bright light. There are more nerves per square inch in your cornea than in any other part of your body!

The autonomic nervous system regulates organ systems through circuits that resemble the reflexes of the somatic nervous system such as the knee jerk reflex. The main difference between the somatic and autonomic systems is in what target tissues are effectors. Somatic responses are solely based on skeletal muscle contraction. The autonomic system, however, targets cardiac and smooth muscle, as well as glandular tissue. Whereas the basic circuit is a reflex arc, there are differences in the structure of those reflexes for the somatic and autonomic systems.

Components of the Reflex Arc

There are typically three neurons in a reflex arc. These are a sensory neuron, which responds to a sensory stimulus (touch, pain, muscle stretch, etc.); an interneuron, which receives a signal if the sensory neuron is stimulated sufficiently; and a motor neuron, which is stimulated by the interneuron and then carries out the action required for the stimulus which initiated the response. The sensory neurons can be oriented externally (i.e., to detect stimuli coming from outside the body), in internally, to detect stimuli originating in the body. An example of the latter would be the stretch receptors that tell your stomach to start contracting harder after a big meal, or the pressure receptors (baroreceptors) in your carotid arteries which detect blood pressure and tell your heart to beat faster or slower, according to the pressure that they are sensing. Similarly, the interneurons fall into two general categories. In long reflexes, the interneuron is in a central nervous system (CNS) structure such as the brain or spinal cord. In short reflexes, the interneuron is located in a peripheral ganglion, bypassing the CNS, as shown in **Figure 29.40**. Finally, the motor neurons which generates the response can also be classified into two general categories. One type of motor neuron innervates a muscle, and stimulates contraction of that muscle, and the subsequent rapid removal of your hand from a hot stove, or a more rapid heartbeat, for example. The second class innervates a gland, and causes secretion of hormones that mediate the appropriate response. An example of the latter would be the nerves which cause the adrenal gland to release adrenaline as part of the fight-or-flight response when you see a lion or tiger or bear coming at you.

Short and Long Reflexes

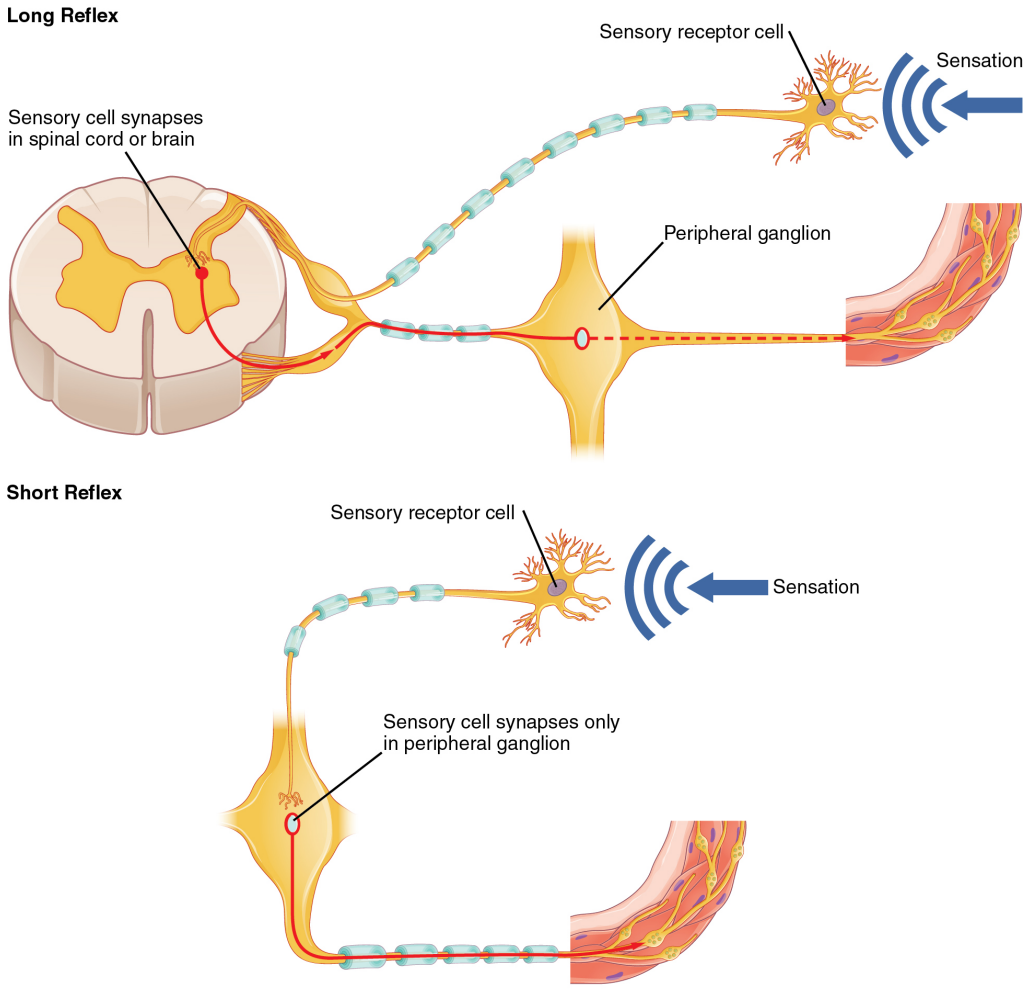


Figure 29.40 Short and Long Reflexes Sensory input can stimulate either a short or a long reflex. A sensory neuron can project to the CNS or to an autonomic ganglion. The short reflex involves the direct stimulation of a postganglionic fiber by the sensory neuron, whereas the long reflex involves integration in the spinal cord or brain.

Homeostasis

As we come to the end of this module, it is appropriate to revisit the concept of homeostasis as a unifying theme in biology. The reflexes described above are just one example homeostatic mechanisms that allow organisms to maintain the parameters of their internal environment about an optimal setting. The organ systems described in this section also all have many examples of homeostatic mechanisms that allow those organ systems to maintain optimal levels of other parameters, such as oxygen levels in the blood, or the pH of the contents of the duodenum. All of these organ systems work together for the benefit, survival and reproduction of the organism. The organ systems are highly interconnected as well. The circulatory and respiratory systems must coordinate their functions, since the function of one is to circulate the needed gases and waste products that come in and go out via the other system. The osmoregulatory system also has to work hand-in-hand with the circulatory and respiratory system to maintain the content of the blood at optimal levels, the digestive system has to provide the nutrients for all of these other organs to function, the nervous and endocrine systems have to respond to a variety of external and internal signals, etc. Finally, the organisms that compose populations and communities and ecosystems also seem to interact with the abiotic components (e.g. nutrients) of the ecosystem in a homeostatic fashion; for example, the CO_2 needed by plants is emitted by other organisms, and the level of atmospheric CO_2 has remained relatively constant over many millenia.

But it is becoming ever more clear that the activities of one species on the planet, *Homo sapiens*, is resetting the set points for these homeostatic processes. As we currently burn about a million years worth of past net primary productivity every year, we are seeing the atmospheric CO_2 levels slowly rise, as you learned in an earlier module. The effects of changing the amount of that globally important compound, over a short time span, geologically speaking, are just starting to be

realized. But it is clear that there will be many ramifications, even if we don't yet understand all of them. It is hoped that your newfound understanding of biological interactions and homeostatic mechanisms will enable your generation to devise appropriate strategies that will again demonstrate that the "sapiens" part of our name is a good descriptor for our species. It will take both the knowledge you have gained in this class, and the wisdom you will gain in all your classes and experiences, to meet the challenges ahead.

ANSWER KEY

<ext:end-of-book-solutions-placeholder></ext:end-of-book-solutions-placeholder>

INDEX

Symbols

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

ATTRIBUTIONS

Collection: **Principles of Biology**
Edited by: Robert Bear, David Rintoul, Bruce Snyder, Martha Smith-Caldas, Christopher Herren, and Eva Horne
Edited by: Robert Bear, David Rintoul, Robert Bear, and David Rintoul
URL: **http://legacy.cnx.org/content/col11569/1.25/**
Copyright: Robert Bear, David Rintoul, Bruce Snyder, Martha Smith-Caldas, Christopher Herren, and Eva Horne
License: **http://creativecommons.org/licenses/by/4.0/**

Module: **Studio Biology - What is it?**
By: David Rintoul and Robert Bear
Edited by: Robert Bear
URL: **http://legacy.cnx.org/content/m47258/1.5/**
Copyright: David Rintoul and Robert Bear
License: **http://creativecommons.org/licenses/by/4.0/**

Module: **Science as a way of knowing**
By: David Rintoul and Robert Bear
URL: **http://legacy.cnx.org/content/m47259/1.10/**
Copyright: David Rintoul and Robert Bear
License: **http://creativecommons.org/licenses/by/4.0/**

Module: **Evolution**
By: David Rintoul and Robert Bear
URL: **http://legacy.cnx.org/content/m47257/1.7/**
Copyright: David Rintoul and Robert Bear
License: **http://creativecommons.org/licenses/by/4.0/**

Module: **Taxonomy and phylogeny**
By: David Rintoul and Robert Bear
Edited by: Robert Bear
Translated by: Robert Bear
URL: **http://legacy.cnx.org/content/m47311/1.7/**
Copyright: David Rintoul and Robert Bear
License: **http://creativecommons.org/licenses/by/4.0/**

Module: **The Scope of Ecology**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m46864/1.7/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: The Scope of Ecology <**http://legacy.cnx.org/content/m44856/1.3**> by OpenStax.

Module: **Ecology of Ecosystems**
By: David Rintoul and Robert Bear
Edited by: Robert Bear
Translated by: Robert Bear
URL: **http://legacy.cnx.org/content/m47376/1.8/**
Copyright: David Rintoul and Robert Bear
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Ecology of Ecosystems <**http://legacy.cnx.org/content/m46868/1.1**> by Robert Bear.

Module: **The Laws of Thermodynamics**
By: Robert Bear and David Rintoul
Edited by: David Rintoul

URL: **http://legacy.cnx.org/content/m46869/1.6/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: The Laws of Thermodynamics <**http://legacy.cnx.org/content/m44424/1.7**> by OpenStax.

Module: **Energy Flow**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m46870/1.8/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Energy Flow through Ecosystems <**http://legacy.cnx.org/content/m44887/1.8**> by OpenStax.

Module: **Biogeochemical Cycles**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m46871/1.13/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Biogeochemical Cycles <**http://legacy.cnx.org/content/m44889/1.6**> by OpenStax.

Module: **Biogeography**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m46872/1.5/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Biogeography <**http://legacy.cnx.org/content/m44857/1.4**> by OpenStax.

Module: **Biomes**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m46873/1.7/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Terrestrial Biomes <**http://legacy.cnx.org/content/m44858/1.4**> by OpenStax.

Module: **Aquatic Biomes**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m46874/1.6/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Aquatic Biomes <**http://legacy.cnx.org/content/m44859/1.4**> by OpenStax.

Module: **Population**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m46877/1.5/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Population Demography <**http://legacy.cnx.org/content/m44864/1.5**> by OpenStax.

Module: **Population Growth**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m46878/1.8/**
Copyright: Robert Bear and David Rintoul

License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Environmental Limits to Population Growth <**http://legacy.cnx.org/content/m44872/1.4**> by OpenStax.

Module: **Population Regulation**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m46879/1.5/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Population Dynamics and Regulation <**http://legacy.cnx.org/content/m44882/1.3**> by OpenStax.

Module: **Human Population Growth**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m46880/1.5/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Human Population Growth <**http://legacy.cnx.org/content/m44875/1.4**> by OpenStax.

Module: **Community Ecology**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m46881/1.8/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Community Ecology <**http://legacy.cnx.org/content/m44878/1.4**> by OpenStax.

Module: **Ecosystem Experimentation and Modeling**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47261/1.4/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/3.0/**
Based on: Ecology of Ecosystems <**http://legacy.cnx.org/content/m44885/1.5**> by OpenStax.

Module: **Nitrogen and Phosphorus Cycles**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47265/1.6/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Biogeochemical Cycles <**http://legacy.cnx.org/content/m44889/1.6**> by OpenStax.

Module: **Freshwater Biomes**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47264/1.4/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/3.0/**
Based on: Aquatic Biomes <**http://legacy.cnx.org/content/m44859/1.4**> by OpenStax.

Module: **Population Growth Curves**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47266/1.6/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Environmental Limits to Population Growth <**http://legacy.cnx.org/content/m44872/1.4**> by OpenStax.

Module: **Introduction to Water Pollution**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47268/1.5/**
Copyright: Robert Bear and David Rintoul

License: http://creativecommons.org/licenses/by/4.0/
Based on: Water Pollution <http://legacy.cnx.org/content/m41441/1.7> by Steve Altaner.

Module: **Introduction to Cells**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: http://legacy.cnx.org/content/m47170/1.6/
Copyright: Robert Bear and David Rintoul
License: http://creativecommons.org/licenses/by/4.0/
Based on: Studying Cells <http://legacy.cnx.org/content/m44405/1.7> by OpenStax.

Module: **Prokaryotic Cells**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: http://legacy.cnx.org/content/m47171/1.4/
Copyright: Robert Bear and David Rintoul
License: http://creativecommons.org/licenses/by/4.0/
Based on: Prokaryotic Cells <http://legacy.cnx.org/content/m44406/1.8> by OpenStax.

Module: **Eukaryotic Cells**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: http://legacy.cnx.org/content/m47172/1.6/
Copyright: Robert Bear and David Rintoul
License: http://creativecommons.org/licenses/by/4.0/
Based on: Eukaryotic Cells <http://legacy.cnx.org/content/m44407/1.11> by OpenStax.

Module: **Protists**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: http://legacy.cnx.org/content/m47173/1.3/
Copyright: Robert Bear and David Rintoul
License: http://creativecommons.org/licenses/by/4.0/
Based on: Characteristics of Protists <http://legacy.cnx.org/content/m44616/1.5> by OpenStax.

Module: **Fungi**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: http://legacy.cnx.org/content/m47174/1.6/
Copyright: Robert Bear and David Rintoul
License: http://creativecommons.org/licenses/by/4.0/
Based on: Fungi <http://legacy.cnx.org/content/m45515/1.3> by OpenStax.

Module: **Eukaryotic Origins**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: http://legacy.cnx.org/content/m47175/1.4/
Copyright: Robert Bear and David Rintoul
License: http://creativecommons.org/licenses/by/3.0/
Based on: Eukaryotic Origins <http://legacy.cnx.org/content/m44614/1.5> by OpenStax.

Module: **Atoms, Isotopes, Ions, and Molecules: The Building Blocks**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: http://legacy.cnx.org/content/m47181/1.9/
Copyright: Robert Bear and David Rintoul
License: http://creativecommons.org/licenses/by/4.0/
Based on: Atoms, Isotopes, Ions, and Molecules: The Building Blocks <http://legacy.cnx.org/content/m44390/1.6>

by OpenStax.

Module: **Water: the Molecule of Life**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47182/1.3/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/3.0/**
Based on: Water <**http://legacy.cnx.org/content/m44392/1.8**> by OpenStax.

Module: **Introduction to Biological Molecules**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m47183/1.3/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**

Module: **Chemical Reactions of Biological Macromolecules**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m47184/1.3/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Synthesis of Biological Macromolecules <**http://legacy.cnx.org/content/m44397/1.6**> by OpenStax.

Module: **Carbohydrates**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m47185/1.6/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Carbohydrates <**http://legacy.cnx.org/content/m44400/1.6**> by OpenStax.

Module: **Lipids**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m47186/1.5/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Lipids <**http://legacy.cnx.org/content/m44401/1.8**> by OpenStax.

Module: **Components and Structure of Plasmal Membranes**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m47190/1.4/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Components and Structure <**http://legacy.cnx.org/content/m44416/1.8**> by OpenStax.

Module: **Proteins**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m47196/1.6/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Proteins <**http://legacy.cnx.org/content/m44402/1.7**> by OpenStax.

Module: **Enzymes**
By: Robert Bear and David Rintoul
Edited by: David Rintoul

URL: **http://legacy.cnx.org/content/m47195/1.7/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Enzymes <**http://legacy.cnx.org/content/m44429/1.6**> by OpenStax.

Module: **Buffers and Enzymes**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m47198/1.5/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Water <**http://legacy.cnx.org/content/m44392/1.8**> by OpenStax.

Module: **Nucleic Acids and Nucleotides**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m47202/1.7/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Nucleic Acids <**http://legacy.cnx.org/content/m44403/1.6**> by OpenStax.

Module: **Cell Reproduction**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m47203/1.4/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/3.0/**
Based on: Cell Division <**http://legacy.cnx.org/content/m44459/1.7**> by OpenStax.

Module: **DNA Replication**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m47204/1.3/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/3.0/**
Based on: DNA Replication <**http://legacy.cnx.org/content/m45475/1.3**> by OpenStax.

Module: **Prokaryotic Cell Division**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m47205/1.4/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/3.0/**
Based on: Prokaryotic Cell Division <**http://legacy.cnx.org/content/m45463/1.4**> by OpenStax.

Module: **Eukaryotic Cell Cycle**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m47206/1.6/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: The Cell Cycle <**http://legacy.cnx.org/content/m45461/1.4**> by OpenStax.

Module: **Cancer and the Cell Cycle**
By: Robert Bear and David Rintoul
Edited by: David Rintoul
URL: **http://legacy.cnx.org/content/m47201/1.4/**
Copyright: Robert Bear and David Rintoul

License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Cancer and the Cell Cycle <**http://legacy.cnx.org/content/m45462/1.3**> by OpenStax.

Module: **Transcription**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47291/1.5/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Transcription <**http://legacy.cnx.org/content/m45476/1.3**> by OpenStax.

Module: **Translation**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47292/1.4/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Translation <**http://legacy.cnx.org/content/m45479/1.5**> by OpenStax.

Module: **Connection Between DNA and Phenotype**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47293/1.6/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**

Module: **Sexual Reproduction**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47294/1.4/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Sexual Reproduction <**http://legacy.cnx.org/content/m45465/1.3**> by OpenStax.

Module: **Meiosis**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47295/1.10/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Meiosis <**http://legacy.cnx.org/content/m45466/1.3**> by OpenStax.

Module: **Mendel's Experiments**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47296/1.7/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Mendel's Experiments <**http://legacy.cnx.org/content/m45469/1.3**> by OpenStax.

Module: **Laws of Inheritance**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47297/1.4/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Laws of Inheritance <**http://legacy.cnx.org/content/m45470/1.3**> by OpenStax.

Module: **Extensions of the Laws of Inheritance**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47304/1.4/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Extensions of the Laws of Inheritance <**http://legacy.cnx.org/content/m45471/1.5**> by OpenStax.

Module: **Population Evolution**

By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47305/1.6/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Population Evolution <**http://legacy.cnx.org/content/m44582/1.5**> by OpenStax.

Module: **Population Genetics**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47306/1.5/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Population Genetics <**http://legacy.cnx.org/content/m44584/1.5**> by OpenStax.

Module: **Formation of New Species**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47307/1.7/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Formation of New Species <**http://legacy.cnx.org/content/m44574/1.5**> by OpenStax.

Module: **Energy and Metabolism**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47324/1.4/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Energy and Metabolism <**http://legacy.cnx.org/content/m44422/1.8**> by OpenStax.

Module: **Thermodynamics**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47325/1.3/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: The Laws of Thermodynamics <**http://legacy.cnx.org/content/m44424/1.7**> by OpenStax.

Module: **Potential, Kinetic, and Free Energy**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47326/1.5/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Potential, Kinetic, Free, and Activation Energy <**http://legacy.cnx.org/content/m44425/1.6**> by OpenStax.

Module: **Energy in Living Systems**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47327/1.7/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Energy in Living Systems <**http://legacy.cnx.org/content/m44431/1.8**> by OpenStax.

Module: **Biological Membranes and Passive Transport**
Used here as: Cell Membranes and Passive Transport
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47328/1.7/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Passive Transport <**http://legacy.cnx.org/content/m44417/1.10**> by OpenStax.

Module: **Energy Requiring Transport**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47329/1.6/**

Based on: Active Transport <http://legacy.cnx.org/content/m45435/1.4> by OpenStax.

Module: **Overview of Photosynthesis**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47330/1.6/

Based on: Overview of Photosynthesis <http://legacy.cnx.org/content/m45448/1.5> by OpenStax.

Module: **The Light-Dependent Reactions**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47331/1.9/

Based on: The Light-Dependent Reactions of Photosynthesis <http://legacy.cnx.org/content/m44448/1.6> by OpenStax.

Module: **Calvin Cycle**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47332/1.8/

Based on: Using Light Energy to Make Organic Molecules <http://legacy.cnx.org/content/m44449/1.6> by OpenStax.

Module: **Overview of Cellular Respiration**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m49546/1.9/

Module: **Glycolysis**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47333/1.5/

Based on: Glycolysis <http://legacy.cnx.org/content/m44432/1.6> by OpenStax.

Module: **Oxidation of Pyruvate and the Krebs Cycle**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47334/1.8/

Based on: Oxidation of Pyruvate and the Citric Acid Cycle <http://legacy.cnx.org/content/m44433/1.7> by OpenStax.

Module: **Oxidative Phosphorylation**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47335/1.9/

Based on: Oxidative Phosphorylation <http://legacy.cnx.org/content/m44437/1.7> by OpenStax.

Module: **Metabolism Without Oxygen**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47336/1.6/

License: http://creativecommons.org/licenses/by/4.0/
Based on: Metabolism without Oxygen <http://legacy.cnx.org/content/m44444/1.6> by OpenStax.

Module: **Connections of Carbohydrate, Protein, and Lipid Metabolic Pathways**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47337/1.7/
Copyright: Robert Bear and David Rintoul
License: http://creativecommons.org/licenses/by/4.0/
Based on: Connections of Carbohydrate, Protein, and Lipid Metabolic Pathways <http://legacy.cnx.org/content/m44441/1.7> by OpenStax.

Module: **Prokaryotes**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47338/1.7/
Copyright: Robert Bear and David Rintoul
License: http://creativecommons.org/licenses/by/4.0/
Based on: Prokaryotic Diversity <http://legacy.cnx.org/content/m44603/1.4> by OpenStax.

Module: **Kingdom Fungi**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47343/1.7/
Copyright: Robert Bear and David Rintoul
License: http://creativecommons.org/licenses/by/4.0/
Based on: Fungi <http://legacy.cnx.org/content/m45515/1.3> by OpenStax.

Module: **Lichens, Protists and Green Algae**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47381/1.13/
Copyright: Robert Bear and David Rintoul
License: http://creativecommons.org/licenses/by/4.0/
Based on: Groups of Protists <http://legacy.cnx.org/content/m44617/1.7> by OpenStax.

Module: **Early Plant Life**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47386/1.9/
Copyright: Robert Bear and David Rintoul
License: http://creativecommons.org/licenses/by/4.0/
Based on: Early Plant Life <http://legacy.cnx.org/content/m44635/1.6> by OpenStax.

Module: **Bryophytes**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47388/1.6/
Copyright: Robert Bear and David Rintoul
License: http://creativecommons.org/licenses/by/4.0/
Based on: Bryophytes <http://legacy.cnx.org/content/m44638/1.6> by OpenStax.

Module: **Seedless Vascular Plants**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47389/1.8/
Copyright: Robert Bear and David Rintoul
License: http://creativecommons.org/licenses/by/4.0/
Based on: Seedless Vascular Plants <http://legacy.cnx.org/content/m44640/1.4> by OpenStax.

Module: **Evolution of Seed Plants**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47399/1.6/
Copyright: Robert Bear and David Rintoul
License: http://creativecommons.org/licenses/by/4.0/
Based on: Evolution of Seed Plants <http://legacy.cnx.org/content/m44653/1.6> by OpenStax.

Module: **Gymnosperms**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47390/1.5/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Gymnosperms <**http://legacy.cnx.org/content/m44648/1.4**> by OpenStax.

Module: **Angiosperms**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47391/1.6/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Angiosperms <**http://legacy.cnx.org/content/m44650/1.4**> by OpenStax.

Module: **Asexual Reproduction**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47393/1.5/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Asexual Reproduction <**http://legacy.cnx.org/content/m44725/1.5**> by OpenStax.

Module: **The Plant Body**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47400/1.4/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: The Plant Body <**http://legacy.cnx.org/content/m44700/1.7**> by OpenStax.

Module: **Plant Sensory Systems and Responses**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47403/1.6/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Plant Sensory Systems and Responses <**http://legacy.cnx.org/content/m44711/1.6**> by OpenStax.

Module: **Reproductive Development and Structure**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47401/1.7/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Reproductive Development and Structure <**http://legacy.cnx.org/content/m44722/1.9**> by OpenStax.

Module: **Pollination and Fertilization**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47402/1.8/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Pollination and Fertilization <**http://legacy.cnx.org/content/m44723/1.5**> by OpenStax.

Module: **Root and Leaf Structure**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47405/1.7/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Roots <**http://legacy.cnx.org/content/m44704/1.4**> by OpenStax.

Module: **Transport of Water and Solutes in Plants**
By: Robert Bear and David Rintoul
Edited by: David Rintoul

URL: **http://legacy.cnx.org/content/m47404/1.5/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Transport of Water and Solutes in Plants <**http://legacy.cnx.org/content/m44708/1.5**> by OpenStax.

Module: **Nutritional Requirements of Plants**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47407/1.3/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Nutritional Requirements of Plants <**http://legacy.cnx.org/content/m44714/1.6**> by OpenStax.

Module: **Nutritional Adaptations of Plants**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47408/1.10/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Nutritional Adaptations of Plants <**http://legacy.cnx.org/content/m44718/1.6**> by OpenStax.

Module: **Photosynthetic Pathways**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47583/1.9/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**

Module: **Climate and the Effects of Global Climate Change**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47410/1.6/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Climate and the Effects of Global Climate Change <**http://legacy.cnx.org/content/m44861/1.3**> by OpenStax.

Module: **Human Population Continues to Grow**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47409/1.3/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Human Population Growth <**http://legacy.cnx.org/content/m44875/1.4**> by OpenStax.

Module: **Features of the Animal Kingdom**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47416/1.10/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Features of the Animal Kingdom <**http://legacy.cnx.org/content/m45523/1.2**> by OpenStax.

Module: **Animal Tissue Types**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47417/1.6/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Animal Primary Tissues <**http://legacy.cnx.org/content/m44731/1.4**> by OpenStax.

Module: **Sponges and Cnidarians**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47418/1.4/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**

Based on: Sponges and Cnidarians <http://legacy.cnx.org/content/m45524/1.3> by OpenStax.

Module: **Flatworms, Nematodes, and Arthropods**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47419/1.3/
Copyright: Robert Bear and David Rintoul
License: http://creativecommons.org/licenses/by/4.0/
Based on: Flatworms, Nematodes, and Arthropods <http://legacy.cnx.org/content/m45525/1.2> by OpenStax.

Module: **Mollusks and Annelids**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47420/1.5/
Copyright: Robert Bear and David Rintoul
License: http://creativecommons.org/licenses/by/4.0/
Based on: Mollusks and Annelids <http://legacy.cnx.org/content/m45526/1.3> by OpenStax.

Module: **Echinoderms and Chordates**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47421/1.4/
Copyright: Robert Bear and David Rintoul
License: http://creativecommons.org/licenses/by/4.0/
Based on: Echinoderms and Chordates <http://legacy.cnx.org/content/m45531/1.2> by OpenStax.

Module: **Vertebrates**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47425/1.3/
Copyright: Robert Bear and David Rintoul
License: http://creativecommons.org/licenses/by/4.0/
Based on: Vertebrates <http://legacy.cnx.org/content/m45532/1.3> by OpenStax.

Module: **Homeostasis**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47426/1.4/
Copyright: Robert Bear and David Rintoul
License: http://creativecommons.org/licenses/by/4.0/
Based on: Homeostasis <http://legacy.cnx.org/content/m44733/1.7> by OpenStax.

Module: **Digestive Systems**
By: Robert Bear, David Rintoul, and Eva Horne
URL: http://legacy.cnx.org/content/m47427/1.8/
Copyright: Robert Bear, David Rintoul, and Eva Horne
License: http://creativecommons.org/licenses/by/4.0/
Based on: Digestive Systems <http://legacy.cnx.org/content/m44736/1.6> by OpenStax.

Module: **Digestive System Processes**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47428/1.7/
Copyright: Robert Bear and David Rintoul
License: http://creativecommons.org/licenses/by/4.0/
Based on: Digestive System Processes <http://legacy.cnx.org/content/m44742/1.4> by OpenStax.

Module: **Digestive System Regulation**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47429/1.4/
Copyright: Robert Bear and David Rintoul
License: http://creativecommons.org/licenses/by/4.0/
Based on: Digestive System Regulation <http://legacy.cnx.org/content/m44744/1.4> by OpenStax.

Module: **Nutrition**

By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47430/1.3/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/3.0/**
Based on: Nutrition and Energy Production <**http://legacy.cnx.org/content/m44740/1.5**> by OpenStax.

Module: **Viruses**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47432/1.10/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Viruses <**http://legacy.cnx.org/content/m45541/1.3**> by OpenStax.

Module: **The Circulatory System**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47435/1.5/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Overview of the Circulatory System <**http://legacy.cnx.org/content/m44801/1.4**> by OpenStax.

Module: **Innate Immunity**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47433/1.4/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Innate Immunity <**http://legacy.cnx.org/content/m45542/1.3**> by OpenStax.

Module: **Adaptive Immunity**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47434/1.4/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Adaptive Immunity <**http://legacy.cnx.org/content/m45543/1.3**> by OpenStax.

Module: **Systems of Gas Exchange**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47436/1.5/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Systems of Gas Exchange <**http://legacy.cnx.org/content/m44792/1.7**> by OpenStax.

Module: **Endocrine System**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47442/1.5/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: Types of Hormones <**http://legacy.cnx.org/content/m44766/1.6**> by OpenStax.

Module: **How Animals Reproduce**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47444/1.4/**
Copyright: Robert Bear and David Rintoul
License: **http://creativecommons.org/licenses/by/4.0/**
Based on: How Animals Reproduce <**http://legacy.cnx.org/content/m45547/1.3**> by OpenStax.

Module: **Urinary System**
By: Robert Bear and David Rintoul
URL: **http://legacy.cnx.org/content/m47445/1.4/**
Copyright: Robert Bear and David Rintoul

Based on: Osmoregulation and Osmotic Balance <http://legacy.cnx.org/content/m44808/1.6> by OpenStax.

Module: **Nervous System**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47519/1.7/

Based on: Neurons and Glial Cells <http://legacy.cnx.org/content/m44747/1.3> by OpenStax.

Module: **Sensory Systems**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47520/1.5/

Based on: Sensory Processes <http://legacy.cnx.org/content/m44754/1.6> by OpenStax.

Module: **Musculoskeletal System**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47521/1.7/

Based on: Types of Skeletal Systems <http://legacy.cnx.org/content/m44781/1.3> by OpenStax.

Module: **Reflexes and Homeostasis**
By: Robert Bear and David Rintoul
URL: http://legacy.cnx.org/content/m47523/1.5/

Based on: Autonomic Reflexes and Homeostasis <http://legacy.cnx.org/content/m46579/1.2> by OpenStax.

ABOUT CONNEXIONS

Since 1999, Connexions has been pioneering a global system where anyone can create course materials and make them fully accessible and easily reusable free of charge. We are a Web-based authoring, teaching and learning environment open to anyone interested in education, including students, teachers, professors and lifelong learners. We connect ideas and facilitate educational communities. Connexions's modular, interactive courses are in use worldwide by universities, community colleges, K-12 schools, distance learners, and lifelong learners. Connexions materials are in many languages, including English, Spanish, Chinese, Japanese, Italian, Vietnamese, French, Portuguese, and Thai.