Concepts of Biology - 1st Canadian Edition

Concepts of Biology - 1st Canadian Edition

Charles Molnar and Jane Gair

Connie Rye; Yael Avissar; Vladimir Jurukovski; Samantha Fowler; Robert Wise; Rebecca Roush; Jung Choi; Jean DeSaix; James Wise; and Michelle Nakano

BCCAMPUS
VICTORIA, B.C.

This textbook was adapted from Concepts of Biology © 2013 by OpenStax, which is under a CC BY 4.0 Licence. The book was adapted by Charles Molnar and Jane Gair as follows:

1. Remixed *Concepts of Biology* from 6 units into 4 units.
2. Removed the original Unit 4: Evolution and the Diversity of Life from *Concepts of Biology.*
3. Remixed the original Unit 5: Animal Structure and Function from *Concepts of Biology* by embedding Chapters 33-43 from *Biology* by OpenStax College.
4. Adapted PowerPoints for each chapter- includes additional notes, images, and embedded videos.
5. Added resources from “Let’s Talk Science” to the end of each PowerPoint.

In 2021, 80 interactive H5P activities by Charles Molnar and Michelle Nakano were added to throughout the text under a CC BY 4.0 Licence.

The CC licence permits you to retain, reuse, copy, redistribute, and revise this book—in whole or in part—for free providing the authors are attributed as follows:

> *Concepts of Biology – 1st Canadian Edition* by Charles Molnar and Jane Gair is used under a CC BY 4.0 Licence.

If you redistribute all or part of this book, it is recommended the following statement be added to the copyright page so readers can access the original book at no cost:

> Download for free from the B.C. Open Textbook Collection.

Sample APA-style citation:

This textbook can be referenced. In APA citation style, it would appear as follows:

Molnar, C., & Gair, J. (2015). *Concepts of Biology – 1st Canadian Edition*. BCcampus. Retrieved from https://opentextbc.ca/biology/

Cover image attribution:

Ebook ISBN: 978-1-989623-99-2

Print ISBN: 978-1-989623-98-5

Visit BCcampus Open Education to learn about open education in British Columbia.

This book was produced with Pressbooks (https://pressbooks.com) and rendered with Prince.

Contents

About BCcampus Open Education

Concepts of Biology—1st Canadian Edition was adapted by Charles Molnar and Jane Gair from the OpenStax textbook *Concepts of Biology*. For information about what was changed, refer to the copyright statement (if using the webbook, that is located at the bottom of the home page). In 2021, this book was updated to include 80 interactive H5P activities.

This adaptation and the creation of the H5P activities was funded by BCcampus Open Education. BCcampus Open Education began in 2012 as the B.C. Open Textbook Project with the goal of making post-secondary education in British Columbia more accessible by reducing students' costs through the use of open textbooks and other OER. BCcampus supports the post-secondary institutions of British Columbia as they adapt and evolve their teaching and learning practices to enable powerful learning opportunities for the students of B.C. BCcampus Open Education is funded by the British Columbia Ministry of Advanced Education and Skills Training, and the Hewlett Foundation.

Open educational resources (OER) are teaching, learning, and research resources that, through permissions granted by the copyright holder, allow others to use, distribute, keep, or make changes to them. Our open textbooks are openly licensed using a Creative Commons licence, and are offered in various e-book formats free of charge, or as printed books that are available at cost.

For more information about open education in British Columbia, please visit the BCcampus Open Education website. If you are an instructor who is using this book for a course, please fill out our Adoption of an Open Textbook form.

Preface to the original textbook, by OpenStax College

Concepts of Biology is intended for the introductory biology course for non-science majors taught at most two- and four-year colleges. The scope, sequence, and level of the program are designed to match typical course syllabi. This text includes interesting features that make connections between scientific concepts and the everyday world of students. *Concepts of Biology* conveys the major themes of biology, such as a foundation in evolution, and features a rich and engaging art program.

Welcome to *Concepts of Biology*, an OpenStax College resource. This textbook has been created with several goals in mind: accessibility, customization, and student engagement—all while encouraging students toward high levels of academic scholarship. Instructors and students alike will find that this textbook offers a strong introduction to biology in an accessible format.

About OpenStax College

OpenStax College is a non-profit organization committed to improving student access to quality learning materials. Our free textbooks are developed and peer-reviewed by educators to ensure they are readable, accurate, and meet the scope and sequence requirements of today's college courses. Unlike traditional textbooks, OpenStax College resources live online and are owned by the community of educators using them. Through our partnerships with companies and foundations committed to reducing costs for students, OpenStax College is working to improve access to higher education for all. OpenStax College is an initiative of Rice University and is made possible through the generous support of several philanthropic foundations.

About OpenStax College's Resources

OpenStax College resources provide quality academic instruction. Three key features set our materials apart from others: they can be customized by instructors for each class, they are a "living" resource that grows online through contributions from science educators, and they are available free or for minimal cost.

Customization

OpenStax College learning resources are designed to be customized for each course. Our textbooks provide a solid foundation on which instructors can build, and our resources are conceived and written with flexibility in mind. Instructors can select the sections most relevant to their curricula and create a textbook that speaks directly to the needs of their classes and student body. Teachers are encouraged to expand on existing examples by adding unique context via geographically localized applications and topical connections.

Concepts of Biology can be easily customized using our online platform. Simply select the content most

relevant to your syllabus and create a textbook that speaks directly to the needs of your class. *Concepts of Biology* is organized as a collection of sections that can be rearranged, modified, and enhanced through localized examples or to incorporate a specific theme of your course. This customization feature will help bring biology to life for your students and will ensure that your textbook truly reflects the goals of your course.

Curation

To broaden access and encourage community curation, *Concepts of Biology* is "open source" licensed under a Creative Commons Attribution (CC-BY) license. The scientific community is invited to submit examples, emerging research, and other feedback to enhance and strengthen the material and keep it current and relevant for today's students. Submit your suggestions to info@openstaxcollege.org, and check in on edition status, alternate versions, errata, and news on the StaxDash at http://openstaxcollege.org.

Cost

Our textbooks are available for free online, and in low-cost print and e-book editions.

About *Concepts of Biology*

Concepts of Biology is designed for the single-semester introduction to biology course for non-science majors, which for many students is their only college-level science course. As such, this course represents an important opportunity for students to develop the necessary knowledge, tools, and skills to make informed decisions as they continue with their lives. Rather than being mired down with facts and vocabulary, the typical non-science major student needs information presented in a way that is easy to read and understand. Even more importantly, the content should be meaningful. Students do much better when they understand why biology is relevant to their everyday lives. For these reasons, *Concepts of Biology* is grounded on an evolutionary basis and includes exciting features that highlight careers in the biological sciences and everyday applications of the concepts at hand. We also strive to show the interconnectedness of topics within this extremely broad discipline. In order to meet the needs of today's instructors and students, we maintain the overall organization and coverage found in most syllabi for this course. A strength of *Concepts of Biology* is that instructors can customize the book, adapting it to the approach that works best in their classroom. *Concepts of Biology* also includes an innovative art program that incorporates critical thinking and clicker questions to help students understand—and apply—key concepts.

Coverage and Scope

Our *Concepts of Biology* textbook adheres to the scope and sequence of most one-semester non-majors courses nationwide. We also strive to make biology, as a discipline, interesting and accessible to students. In addition to a comprehensive coverage of core concepts and foundational research, we have incorporated features that draw learners into the discipline in meaningful ways. Our scope of content was developed after surveying over a hundred biology professors and listening to their coverage needs.

We provide a thorough treatment of biology's fundamental concepts with a scope that is manageable for instructors and students alike.

- Unit 1: **The Cellular Foundation of Life**. Our opening unit introduces students to the sciences, including the process of science and the underlying concepts from the physical sciences that provide a framework within which learners comprehend biological processes. Additionally, students will gain solid understanding of the structures, functions, and processes of the most basic unit of life: the cell.
- Unit 2: **Cell Division and Genetics**. Our genetics unit takes learners from the foundations of cellular reproduction to the experiments that revealed the basis of genetics and laws of inheritance.
- Unit 3: **Molecular Biology and Biotechnology**. Students will learn the intricacies of DNA, protein synthesis, and gene regulation and current applications of biotechnology and genomics.
- Unit 4: **Evolution and the Diversity of Life**. The core concepts of evolution are discussed in this unit with examples illustrating evolutionary processes. Additionally, the evolutionary basis of biology reappears throughout the textbook in general discussion and is reinforced through special call-out features highlighting specific evolution-based topics. The diversity of life is explored with detailed study of various organisms and discussion of emerging phylogenetic relationships between and among bacteria, protist kingdoms, fungi, plants, and animals.
- Unit 5: **Animal Structure and Function**. An introduction to the form and function of the animal body is followed by chapters on the immune system and animal development. This unit touches on the biology of all organisms while maintaining an engaging focus on human anatomy and physiology that helps students connect to the topics.
- Unit 6: **Ecology**. Ecological concepts are broadly covered in this unit, with features highlighting localized, real-world issues of conservation and biodiversity.

Pedagogical Foundation and Features

Because of the impact science has on students and society, an important goal of science education is to achieve a scientifically literate population that consistently makes informed decisions. Scientific literacy transcends a basic understanding of scientific principles and processes to include the ability to make sense of the myriad instances where people encounter science in day-to-day life. Thus, a scientifically literate person is one who uses science content knowledge to make informed decisions, either personally or socially, about topics or issues that have a connection with science. *Concepts of Biology* is grounded on a solid scientific base and designed to promote scientific literacy. Throughout the text, you will find features that engage the students in scientific inquiry by taking selected topics a step further.

- **Evolution in Action** features uphold the importance of evolution to all biological study through discussions like "Global Decline of Coral Reefs" and "The Red Queen Hypothesis."
- **Career in Action** features present information on a variety of careers in the biological sciences, introducing students to the educational requirements and day-to-day work life of a

variety of professions, such as forensic scientists, registered dietitians, and biogeographers.

- **Biology in Action** features tie biological concepts to emerging issues and discuss science in terms of everyday life. Topics include “Invasive Species” and “Photosynthesis at the Grocery Store.”

Art and Animations that Engage

Our art program takes a straightforward approach designed to help students learn the concepts of biology through simple, effective illustrations, photos, and micrographs. *Concepts of Biology* also incorporates links to relevant animations and interactive exercises that help bring biology to life for students.

- **Concepts in Action** features direct students to online interactive exercises and animations to add a fuller context and examples to core content.

About Our Team

Concepts of Biology would not be possible if not for the tremendous contributions of the authors and community reviewing team

Senior Contributors

- Samantha Fowler, Clayton State University
- Rebecca Roush, Sandhills Community College
- James Wise, Hampton University

Faculty Contributors and Reviewers

- Mark Belk, Brigham Young University
- Lisa Boggs, Southwestern Oklahoma State University
- Sherryl Broverman, Duke University
- David Byres, Florida State College at Jacksonville
- Aaron Cassill, The University of Texas at San Antonio
- Karen Champ, College of Central Florida
- Sue Chaplin, University of St. Thomas
- Diane Day, Clayton State University
- Jean DeSaix, University of North Carolina at Chapel Hill
- David Hunnicutt, St. Norbert College
- Barbara Kuehner, Hawaii Community College
- Brenda Leady, University of Toledo
- Bernie Marcus, Genesee Community College
- Flora Mhlanga, Lipscomb University
- Madeline Mignone, Dominican College
- Elizabeth Nash, Long Beach City College
- Mark Newton, San Jose City College
- Diana Oliveras, University of Colorado Boulder
- Ann Paterson, Williams Baptist College
- Joel Piperberg, Millersville University
- Nick Reeves, Mt. San Jacinto College
- Ann Reisenauer, San Jose State University
- Lynn Rumfelt, Gordon College
- Michael Rutledge, Middle Tennessee State University
- Edward Saiff, Ramapo College of New Jersey

- Brian Shmaefsky, Kingwood College
- Gary Shultz, Marshall University
- Donald Slish, SUNY Plattsburgh
- Anh-Hue Tu, Georgia Southwestern State University
- Elena Zoubina, Bridgewater State University

Preface to the 1st Canadian Edition

Preface to the 1st Canadian Edition, by Charles Molnar and Jane Gair, adapters of *Concepts of Biology*

In this survey text, directed at those not majoring in biology, we dispel the assumption that a little learning is a dangerous thing. We hope that by skimming the surface of a very deep subject, biology, we may inspire you to drink more deeply and make more informed choices relating to your health, the environment, politics, and the greatest subject that all of us are entwined in, life itself.

In the adapted textbook, *Concepts of Biology — 1st Canadian Edition*, you will find the following units:

- Unit 1: The Cellular Foundation of Life
- Unit 2: Cell Division and Genetics
- Unit 3: Molecular Biology and Biotechnology
- Unit 4: Animal Structure and Function

Adaptations to the original textbook *Concepts of Biology* by OpenStax College include:

- Remixed *Concepts of Biology* from 6 units into 4 units.
- Removed the original Unit 4: Evolution and the Diversity of Life from *Concepts of Biology.*
- Remixed the original Unit 5: Animal Structure and Function from *Concepts of Biology* by embedding Chapters 33-43 from *Biology* by OpenStax College.
- Adapted PowerPoints for each chapter- includes additional notes, images, and embedded videos.
- Added resources from “Let’s Talk Science” to the end of each PowerPoint.
- Added 80 H5P activities throughout the book.

Thanks to BCcampus and Camosun College for funding and support. We are most grateful to the Let’s Talk Science organization from their trove of science links.

UNIT 1. THE CELLULAR FOUNDATION OF LIFE

Unit 1: The Cellular Foundation of Life includes the following chapters:

Chapter 1: Introduction to Biology

Figure 1.1 This NASA image is a composite of several satellite-based views of Earth. To make the whole-Earth image, NASA scientists combine observations of different parts of the planet. (credit: modification of work by NASA)

Viewed from space, Earth offers few clues about the diversity of life forms that reside there. The first forms of life on Earth are thought to have been microorganisms that existed for billions of years before plants and animals appeared. The mammals, birds, and flowers so familiar to us are all relatively recent, originating 130 to 200 million years ago. Humans have inhabited this planet for only the last 2.5 million years, and only in the last 200,000 years have humans started looking like we do today.

Introduction to Interactive Learning

The goal of interactive learning is to promote engagement with and retention of the concepts and information being studied. The interactive learning activities in this open textbook support self directed practice and are not recorded assessments. Use the interactive learning activities to:

- Search for key points at the start of each chapter
- Interact with videos, and
- Practice review questions within the chapter sections.

Watch an introduction to interactive videos

An interactive H5P element has been excluded from this version of the text. You can view it online here: *https://opentextbc.ca/biology/?p=4308#h5p-1*

Search for Key Points in Chapter 1

Get to know your open textbook.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4308#h5p-2

Media Attributions

- Figure 1.1 by NASA © Public Domain

1.1 Themes and Concepts of Biology

Learning Objectives

By the end of this section, you will be able to:

- Identify and describe the properties of life
- Describe the levels of organization among living things
- List examples of different sub disciplines in biology

Watch a video about Evolution by Natural Selection.

An interactive H5P element has been excluded from this version of the text. You can view it online here: *https://opentextbc.ca/biology/?p=4324#h5p-3*

Biology is the science that studies life. What exactly is life? This may sound like a silly question with an obvious answer, but it is not easy to define life. For example, a branch of biology called virology studies viruses, which exhibit some of the characteristics of living entities but lack others. It turns out that although viruses can attack living organisms, cause diseases, and even reproduce, they do not meet the criteria that biologists use to define life.

From its earliest beginnings, biology has wrestled with four questions: What are the shared properties that make something “alive”? How do those various living things function? When faced with the remarkable diversity of life, how do we organize the different kinds of organisms so that we can better understand them? And, finally—what biologists ultimately seek to understand—how did this diversity arise and how is it continuing? As new organisms are discovered every day, biologists continue to seek answers to these and other questions.

Properties of Life

All groups of living organisms share multiple key characteristics or functions: order, sensitivity or response to stimuli, reproduction, adaptation, growth and development, regulation, homeostasis, and energy processing. When viewed together, these eight characteristics serve to define life.

Order

Organisms are highly organized structures that consist of one or more cells. Even very simple, single-celled organisms are remarkably complex. Inside each cell, atoms make up molecules. These in turn make up cell components or organelles. Multicellular organisms, which may consist of millions of individual cells, have an advantage over single-celled organisms in that their cells can be specialized to perform specific functions, and even sacrificed in certain situations for the good of the organism as a whole. How these specialized cells come together to form organs such as the heart, lung, or skin in organisms like the toad shown in Figure 1. 2 will be discussed later.

Figure 1.2 A toad represents a highly organized structure consisting of cells, tissues, organs, and organ systems.

Sensitivity or Response to Stimuli

Organisms respond to diverse stimuli. For example, plants can bend toward a source of light or respond to touch. Even tiny bacteria can move toward or away from chemicals (a process called chemotaxis) or light (phototaxis). Movement toward a stimulus is considered a positive response, while movement away from a stimulus is considered a negative response.

Figure 1.3 The leaves of this sensitive plant (Mimosa pudica) will instantly droop and fold when touched. After a few minutes, the plant returns to its normal state.

Concept in Action

Watch this video to see how the sensitive plant responds to a touch stimulus.

Reproduction

Single-celled organisms reproduce by first duplicating their DNA, which is the genetic material, and then dividing it equally as the cell prepares to divide to form two new cells. Many multicellular organisms (those made up of more than one cell) produce specialized reproductive cells that will form new individuals. When reproduction occurs, DNA containing genes is passed along to an organism's offspring. These genes are the reason that the offspring will belong to the same species and will have characteristics similar to the parent, such as fur color and blood type.

Adaptation

All living organisms exhibit a "fit" to their environment. Biologists refer to this fit as adaptation and it is a consequence of evolution by natural selection, which operates in every lineage of reproducing

organisms. Examples of adaptations are diverse and unique, from heat-resistant Archaea that live in boiling hot springs to the tongue length of a nectar-feeding moth that matches the size of the flower from which it feeds. All adaptations enhance the reproductive potential of the individual exhibiting them, including their ability to survive to reproduce. Adaptations are not constant. As an environment changes, natural selection causes the characteristics of the individuals in a population to track those changes.

Growth and Development

Organisms grow and develop according to specific instructions coded for by their genes. These genes provide instructions that will direct cellular growth and development, ensuring that a species' young will grow up to exhibit many of the same characteristics as its parents.

Figure 1.4 Although no two look alike, these kittens have inherited genes from both parents and share many of the same characteristics.

Regulation

Even the smallest organisms are complex and require multiple regulatory mechanisms to coordinate internal functions, such as the transport of nutrients, response to stimuli, and coping with environmental stresses. For example, organ systems such as the digestive or circulatory systems perform specific functions like carrying oxygen throughout the body, removing wastes, delivering nutrients to every cell, and cooling the body.

Homeostasis

To function properly, cells require appropriate conditions such as proper temperature, pH, and concentrations of diverse chemicals. These conditions may, however, change from one moment to the next. Organisms are able to maintain internal conditions within a narrow range almost constantly, despite environmental changes, through a process called homeostasis or "steady state"—the ability of an organism to maintain constant internal conditions. For example, many organisms regulate their body temperature in a process known as thermoregulation. Organisms that live in cold climates, such as the

polar bear, have body structures that help them withstand low temperatures and conserve body heat. In hot climates, organisms have methods (such as perspiration in humans or panting in dogs) that help them to shed excess body heat.

Figure 1.5 Polar bears and other mammals living in ice-covered regions maintain their body temperature by generating heat and reducing heat loss through thick fur and a dense layer of fat under their skin.

Energy Processing

All organisms (such as the California condor shown in Figure 1.6) use a source of energy for their metabolic activities. Some organisms capture energy from the sun and convert it into chemical energy in food; others use chemical energy from molecules they take in.

Figure 1.6 A lot of energy is required for a California condor to fly. Chemical energy derived from food is used to power flight. California condors are an endangered species; scientists have strived to place a wing tag on each bird to help them identify and locate each individual bird.

Levels of Organization of Living Things

Living things are highly organized and structured, following a hierarchy on a scale from small to large. The atom is the smallest and most fundamental unit of matter. It consists of a nucleus surrounded by electrons. Atoms form molecules. A **molecule** is a chemical structure consisting of at least two atoms held together by a chemical bond. Many molecules that are biologically important are **macromolecules**, large molecules that are typically formed by combining smaller units called monomers. An example of a macromolecule is deoxyribonucleic acid (DNA), which contains the instructions for the functioning of the organism that contains it.

Figure 1.7 A molecule, like this large DNA molecule, is composed of atoms.

Concept in Action

To see an animation of this DNA molecule, click here.

Some cells contain aggregates of macromolecules surrounded by membranes; these are called organelles. Organelles are small structures that exist within cells and perform specialized functions. All living things are made of cells; the cell itself is the smallest fundamental unit of structure and function in living organisms. (This requirement is why viruses are not considered living: they are not made of cells. To make new viruses, they have to invade and hijack a living cell; only then can they obtain the materials they need to reproduce.) Some organisms consist of a single cell and others are multicellular. Cells are classified as prokaryotic or eukaryotic. Prokaryotes are single-celled organisms that lack organelles surrounded by a membrane and do not have nuclei surrounded by nuclear membranes; in contrast, the cells of eukaryotes do have membrane-bound organelles and nuclei.

In most multicellular organisms, cells combine to make tissues, which are groups of similar cells carrying out the same function. Organs are collections of tissues grouped together based on a common function. Organs are present not only in animals but also in plants. An organ system is a higher level of organization that consists of functionally related organs. For example vertebrate animals have many organ systems, such as the circulatory system that transports blood throughout the body and to and from the lungs; it includes organs such as the heart and blood vessels. Organisms are individual living entities. For example, each tree in a forest is an organism. Single-celled prokaryotes and single-celled eukaryotes are also considered organisms and are typically referred to as microorganisms.

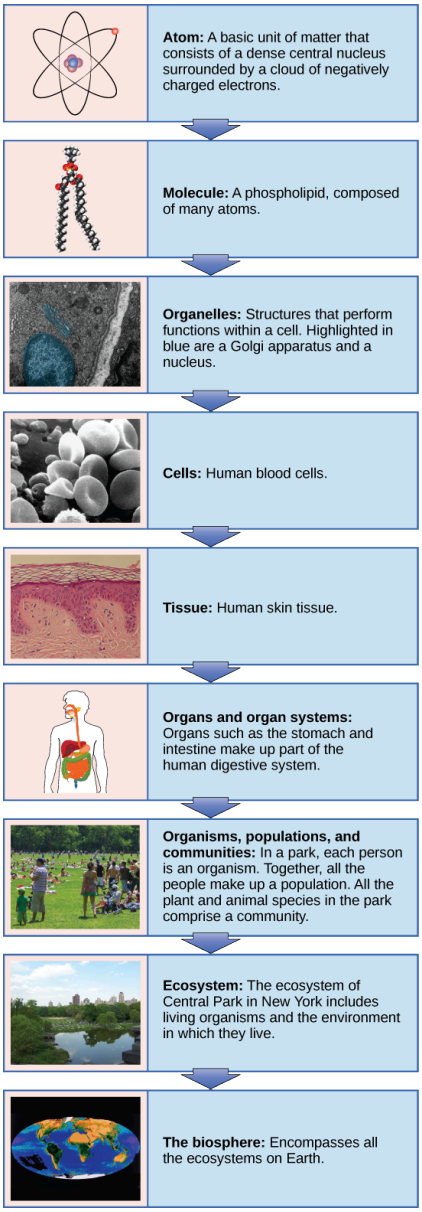


Figure 1.8 From an atom to the entire Earth, biology examines all aspects of life.

Which of the following statements is false?

1. Tissues exist within organs which exist within organ systems.

2. Communities exist within populations which exist within ecosystems.
3. Organelles exist within cells which exist within tissues.
4. Communities exist within ecosystems which exist in the biosphere.

All the individuals of a species living within a specific area are collectively called a population. For example, a forest may include many white pine trees. All of these pine trees represent the population of white pine trees in this forest. Different populations may live in the same specific area. For example, the forest with the pine trees includes populations of flowering plants and also insects and microbial populations. A community is the set of populations inhabiting a particular area. For instance, all of the trees, flowers, insects, and other populations in a forest form the forest's community. The forest itself is an ecosystem. An ecosystem consists of all the living things in a particular area together with the abiotic, or non-living, parts of that environment such as nitrogen in the soil or rainwater. At the highest level of organization, the biosphere is the collection of all ecosystems, and it represents the zones of life on Earth. It includes land, water, and portions of the atmosphere.

The Diversity of Life

The science of biology is very broad in scope because there is a tremendous diversity of life on Earth. The source of this diversity is evolution, the process of gradual change during which new species arise from older species. Evolutionary biologists study the evolution of living things in everything from the microscopic world to ecosystems.

In the 18th century, a scientist named Carl Linnaeus first proposed organizing the known species of organisms into a hierarchical taxonomy. In this system, species that are most similar to each other are put together within a grouping known as a genus. Furthermore, similar genera (the plural of genus) are put together within a family. This grouping continues until all organisms are collected together into groups at the highest level. The current taxonomic system now has eight levels in its hierarchy, from lowest to highest, they are: species, genus, family, order, class, phylum, kingdom, and domain. Thus species are grouped within genera, genera are grouped within families, families are grouped within orders, and so on.

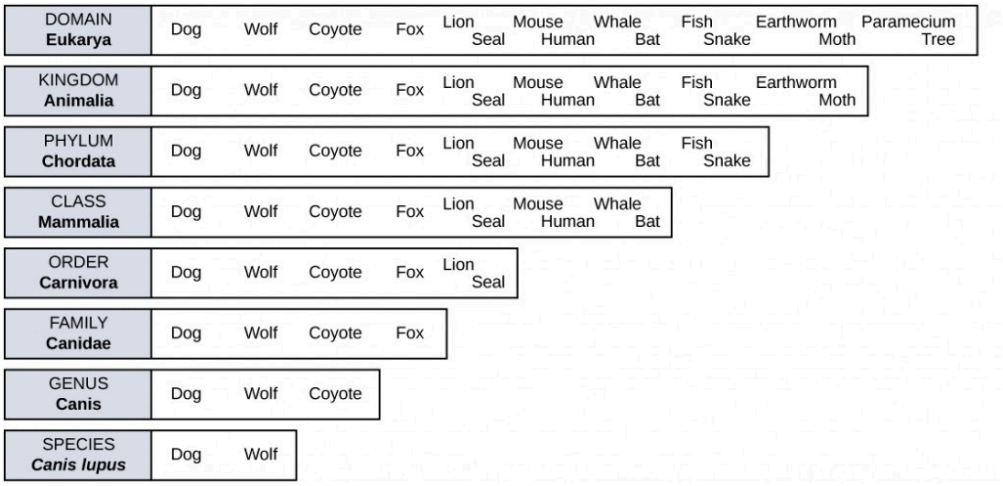


Figure 1.9 This diagram shows the levels of taxonomic hierarchy for a dog, from the broadest category—domain—to the most specific—species.

The highest level, domain, is a relatively new addition to the system since the 1990s. Scientists now recognize three domains of life, the Eukarya, the Archaea, and the Bacteria. The domain Eukarya contains organisms that have cells with nuclei. It includes the kingdoms of fungi, plants, animals, and several kingdoms of protists. The Archaea, are single-celled organisms without nuclei and include many extremophiles that live in harsh environments like hot springs. The Bacteria are another quite different group of single-celled organisms without nuclei. Both the Archaea and the Bacteria are prokaryotes, an informal name for cells without nuclei. The recognition in the 1990s that certain "bacteria," now known as the Archaea, were as different genetically and biochemically from other bacterial cells as they were from eukaryotes, motivated the recommendation to divide life into three domains. This dramatic change in our knowledge of the tree of life demonstrates that classifications are not permanent and will change when new information becomes available.

In addition to the hierarchical taxonomic system, Linnaeus was the first to name organisms using two unique names, now called the binomial naming system. Before Linnaeus, the use of common names to refer to organisms caused confusion because there were regional differences in these common names. Binomial names consist of the genus name (which is capitalized) and the species name (all lower-case). Both names are set in italics when they are printed. Every species is given a unique binomial which is recognized the world over, so that a scientist in any location can know which organism is being referred to. For example, the North American blue jay is known uniquely as *Cyanocitta cristata*. Our own species is *Homo sapiens*.

Figure 1.10 These images represent different domains. The scanning electron micrograph shows (a) bacterial cells belong to the domain Bacteria, while the (b) extremophiles, seen all together as colored mats in this hot spring, belong to domain Archaea. Both the (c) sunflower and (d) lion are part of domain Eukarya.

Evolution in Action

Carl Woese and the Phylogenetic Tree

The evolutionary relationships of various life forms on Earth can be summarized in a phylogenetic tree. A phylogenetic tree is a diagram showing the evolutionary relationships among biological species based on similarities and differences in genetic or physical traits or both. A phylogenetic tree is composed of branch points, or nodes, and branches. The internal nodes represent ancestors and are points in evolution when, based on scientific evidence, an ancestor is thought to have diverged to form two new species. The length of each branch can be considered as estimates of relative time.

In the past, biologists grouped living organisms into five kingdoms: animals, plants, fungi, protists, and bacteria. The pioneering work of American microbiologist Carl Woese in the early 1970s has shown, however, that life on Earth has evolved along three lineages, now called domains—Bacteria, Archaea, and Eukarya. Woese proposed the domain as a new taxonomic level and Archaea as a new domain, to

reflect the new phylogenetic tree. Many organisms belonging to the Archaea domain live under extreme conditions and are called extremophiles. To construct his tree, Woese used genetic relationships rather than similarities based on morphology (shape). Various genes were used in phylogenetic studies. Woese's tree was constructed from comparative sequencing of the genes that are universally distributed, found in some slightly altered form in every organism, conserved (meaning that these genes have remained only slightly changed throughout evolution), and of an appropriate length.

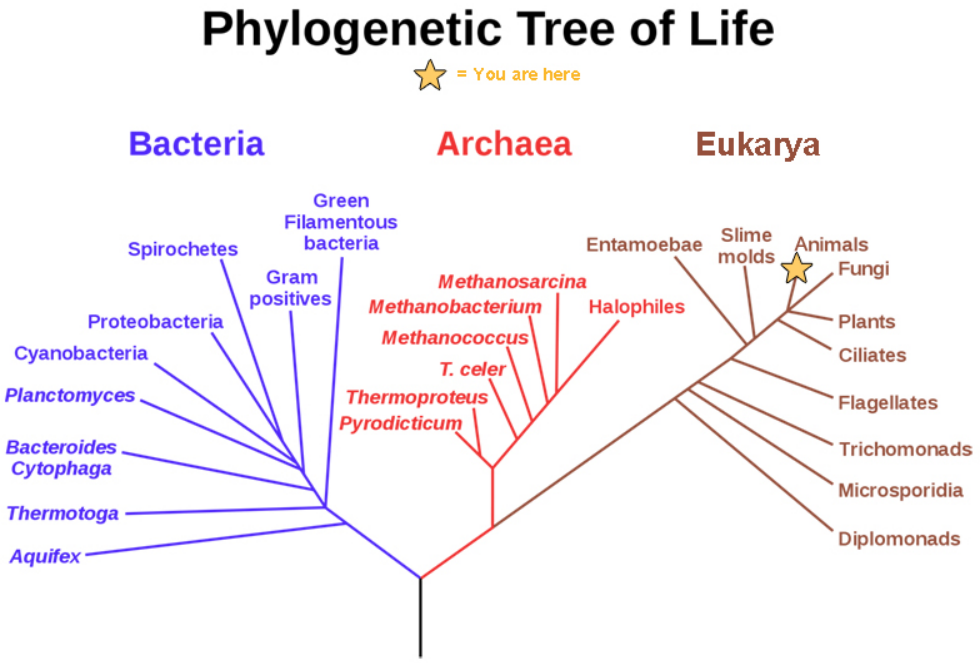


Figure 1.11 This phylogenetic tree was constructed by microbiologist Carl Woese using genetic relationships. The tree shows the separation of living organisms into three domains: Bacteria, Archaea, and Eukarya. Bacteria and Archaea are organisms without a nucleus or other organelles surrounded by a membrane and, therefore, are prokaryotes.

Branches of Biological Study

Watch a video about Science and Medicine

An interactive H5P element has been excluded from this version of the text. You can view it online here: *https://opentextbc.ca/biology/?p=4324#h5p-4*

The scope of biology is broad and therefore contains many branches and sub disciplines. Biologists may pursue one of those sub disciplines and work in a more focused field. For instance, molecular biology studies biological processes at the molecular level, including interactions among molecules such as DNA, RNA, and proteins, as well as the way they are regulated. Microbiology is the study of

the structure and function of microorganisms. It is quite a broad branch itself, and depending on the subject of study, there are also microbial physiologists, ecologists, and geneticists, among others.

Another field of biological study, neurobiology, studies the biology of the nervous system, and although it is considered a branch of biology, it is also recognized as an interdisciplinary field of study known as neuroscience. Because of its interdisciplinary nature, this sub discipline studies different functions of the nervous system using molecular, cellular, developmental, medical, and computational approaches.

Figure 1.12 Researchers work on excavating dinosaur fossils at a site in Castellón, Spain.

Paleontology, another branch of biology, uses fossils to study life's history. Zoology and botany are the study of animals and plants, respectively. Biologists can also specialize as biotechnologists, ecologists, or physiologists, to name just a few areas. Biotechnologists apply the knowledge of biology to create useful products. Ecologists study the interactions of organisms in their environments. Physiologists study the workings of cells, tissues and organs. This is just a small sample of the many fields that biologists can pursue. From our own bodies to the world we live in, discoveries in biology can affect us in very direct and important ways. We depend on these discoveries for our health, our food sources, and the benefits provided by our ecosystem. Because of this, knowledge of biology can benefit us in making decisions in our day-to-day lives.

The development of technology in the twentieth century that continues today, particularly the technology to describe and manipulate the genetic material, DNA, has transformed biology. This transformation will allow biologists to continue to understand the history of life in greater detail, how the human body works, our human origins, and how humans can survive as a species on this planet despite the stresses caused by our increasing numbers. Biologists continue to decipher huge mysteries about life suggesting that we have only begun to understand life on the planet, its history, and our relationship to it. For this and other reasons, the knowledge of biology gained through this textbook and other printed and electronic media should be a benefit in whichever field you enter.

Forensic Scientist

Forensic science is the application of science to answer questions related to the law. Biologists as well as chemists and biochemists can be forensic scientists. Forensic scientists provide scientific evidence for use in courts, and their job involves examining trace material associated with crimes. Interest in forensic science has increased in the last few years, possibly because of popular television shows that feature forensic scientists on the job. Also, the development of molecular techniques and the establishment of DNA databases have updated the types of work that forensic scientists can do. Their job activities are primarily related to crimes against people such as murder, rape, and assault. Their work involves analyzing samples such as hair, blood, and other body fluids and also processing DNA found in many different environments and materials. Forensic scientists also analyze other biological evidence left at crime scenes, such as insect parts or pollen grains. Students who want to pursue careers in forensic science will most likely be required to take chemistry and biology courses as well as some intensive math courses.

Figure 1.13 This forensic scientist works in a DNA extraction room at the U.S. Army Criminal Investigation Laboratory.

Section Summary

Biology is the science of life. All living organisms share several key properties such as order, sensitivity or response to stimuli, reproduction, adaptation, growth and development, regulation, homeostasis, and energy processing. Living things are highly organized following a hierarchy that includes atoms, molecules, organelles, cells, tissues, organs, and organ systems. Organisms, in turn, are grouped as populations, communities, ecosystems, and the biosphere. Evolution is the source of the tremendous biological diversity on Earth today. A diagram called a phylogenetic tree can be used to show evolutionary relationships among organisms. Biology is very broad and includes many branches and sub disciplines. Examples include molecular biology, microbiology, neurobiology, zoology, and botany, among others.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4324#h5p-5

Glossary

atom: a basic unit of matter that cannot be broken down by normal chemical reactions

biology: the study of living organisms and their interactions with one another and their environments

biosphere: a collection of all ecosystems on Earth

cell: the smallest fundamental unit of structure and function in living things

community: a set of populations inhabiting a particular area

ecosystem: all living things in a particular area together with the abiotic, nonliving parts of that environment

eukaryote: an organism with cells that have nuclei and membrane-bound organelles

evolution: the process of gradual change in a population that can also lead to new species arising from older species

homeostasis: the ability of an organism to maintain constant internal conditions

macromolecule: a large molecule typically formed by the joining of smaller molecules

molecule: a chemical structure consisting of at least two atoms held together by a chemical bond

organ: a structure formed of tissues operating together to perform a common function

organ system: the higher level of organization that consists of functionally related organs

organelle: a membrane-bound compartment or sac within a cell

organism: an individual living entity

phylogenetic tree: a diagram showing the evolutionary relationships among biological species based on similarities and differences in genetic or physical traits or both

population: all individuals within a species living within a specific area

prokaryote: a unicellular organism that lacks a nucleus or any other membrane-bound organelle

tissue: a group of similar cells carrying out the same function

Media Attribution

- Figure 1.2 by Ivengo(RUS) © Public Domain
- Figure 1.3 by Alex Lomas © CC BY (Attribution)
- Figure 1.4 by Pieter & Renée Lanser © CC BY (Attribution)
- Figure 1.5 by David © CC BY (Attribution)
- Figure 1.6 by Pacific Southwest Region USFWS © CC BY (Attribution)
- Figure 1.7 by Brian0918 © Public Domain
- Figure 1.8
    - “molecule”: modification of work by Jane Whitney;
    - “organelles”: modification of work by Louisa Howard;
    - “cells”: modification of work by Bruce Wetzel, Harry Schaefer, National Cancer Institute;
    - “tissue”: modification of work by “Kilbad” © Public Domain
    - “organs”: modification of work by Mariana Ruiz Villareal, Joaquim Alves Gaspar;
    - “organisms”: modification of work by Peter Dutton;
    - “ecosystem”: modification of work by “gigi4791″ © CC BY (Attribution)
    - “biosphere”: modification of work by NASA © Public Domain
- Figure 1.10
    - EscherichiaColi NIAID: modification of work by Rocky Mountain Laboratories, NIAID, NIH © Public Domain
    - Extremophiles modification of work by Steve Jurvetson © CC BY (Attribution)
    - Sunflower modification of work by Michael Arrighi
    - Lion modification of work by Frank Vassen © CC BY (Attribution)
- Figure 1.12 by Mario Modesto © Public Domain
- Figure 1.13 by U.S. Army CID Command Public Affairs

1.2 The Process of Science

Learning Objectives

By the end of this section, you will be able to:

- Identify the shared characteristics of the natural sciences
- Understand the process of scientific inquiry
- Compare inductive reasoning with deductive reasoning
- Describe the goals of basic science and applied science

Watch a video about the Scientific Method.

An interactive H5P element has been excluded from this version of the text. You can view it online here: *https://opentextbc.ca/biology/?p=4332#h5p-6*

Figure 1.14 Formerly called blue-green algae, the (a) cyanobacteria seen through a light microscope are some of Earth's oldest life forms. These (b) stromatolites along the shores of Lake Thetis in Western Australia are ancient structures formed by the layering of cyanobacteria in shallow waters.

Like geology, physics, and chemistry, biology is a science that gathers knowledge about the natural world. Specifically, biology is the study of life. The discoveries of biology are made by a community of researchers who work individually and together using agreed-on methods. In this sense, biology, like all sciences is a social enterprise like politics or the arts. The methods of science include careful observation, record keeping, logical and mathematical reasoning, experimentation, and submitting conclusions to the scrutiny of others. Science also requires considerable imagination and creativity; a well-designed experiment is commonly described as elegant, or beautiful. Like politics, science has considerable practical implications and some science is dedicated to practical applications, such as the prevention of disease. Other science proceeds largely motivated by curiosity. Whatever its goal, there is no doubt that science, including biology, has transformed human existence and will continue to do so.

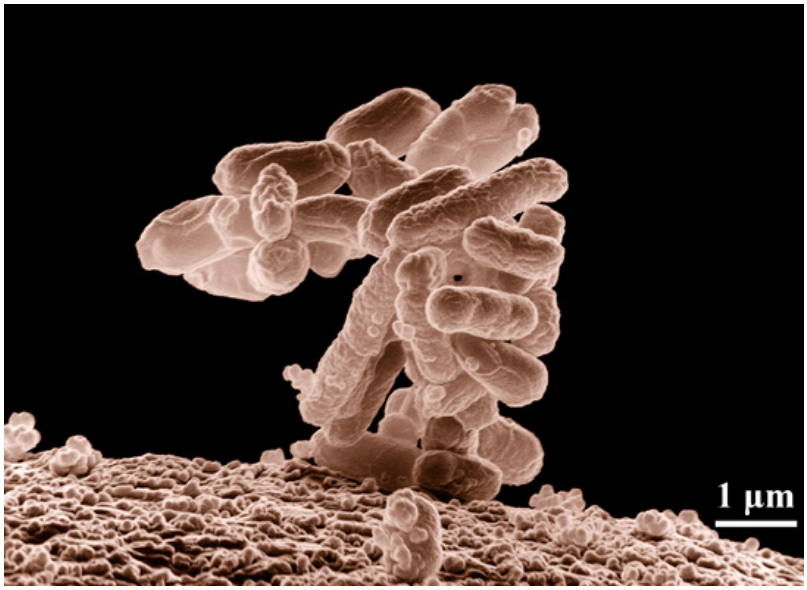


Figure 1.15 Biologists may choose to study Escherichia coli (E. coli), a bacterium that is a normal resident of our digestive tracts but which is also sometimes responsible for disease outbreaks. In this micrograph, the bacterium is visualized using a scanning electron microscope and digital colorization.

The Nature of Science

Watch a video about the reductional approach of western science.

An interactive H5P element has been excluded from this version of the text. You can view it online here: *https://opentextbc.ca/biology/?p=4332#h5p-7*

Biology is a science, but what exactly is science? What does the study of biology share with other

scientific disciplines? Science (from the Latin *scientia,* meaning “knowledge”) can be defined as knowledge about the natural world.

Science is a very specific way of learning, or knowing, about the world. The history of the past 500 years demonstrates that science is a very powerful way of knowing about the world; it is largely responsible for the technological revolutions that have taken place during this time. There are however, areas of knowledge and human experience that the methods of science cannot be applied to. These include such things as answering purely moral questions, aesthetic questions, or what can be generally categorized as spiritual questions. Science has cannot investigate these areas because they are outside the realm of material phenomena, the phenomena of matter and energy, and cannot be observed and measured.

The scientific method is a method of research with defined steps that include experiments and careful observation. The steps of the scientific method will be examined in detail later, but one of the most important aspects of this method is the testing of hypotheses. A hypothesis is a suggested explanation for an event, which can be tested. **Hypotheses**, or **tentative explanations**, are generally produced within the context of a scientific theory. A **scientific theory** is a generally accepted, thoroughly tested and confirmed explanation for a set of observations or phenomena. Scientific theory is the foundation of scientific knowledge. In addition, in many scientific disciplines (less so in biology) there are scientific laws, often expressed in mathematical formulas, which describe how elements of nature will behave under certain specific conditions. There is not an evolution of hypotheses through theories to laws as if they represented some increase in certainty about the world. Hypotheses are the day-to-day material that scientists work with and they are developed within the context of theories. Laws are concise descriptions of parts of the world that are amenable to formulaic or mathematical description.

Natural Sciences

What would you expect to see in a museum of natural sciences? Frogs? Plants? Dinosaur skeletons? Exhibits about how the brain functions? A planetarium? Gems and minerals? Or maybe all of the above? Science includes such diverse fields as astronomy, biology, computer sciences, geology, logic, physics, chemistry, and mathematics. However, those fields of science related to the physical world and its phenomena and processes are considered natural sciences. Thus, a museum of natural sciences might contain any of the items listed above.

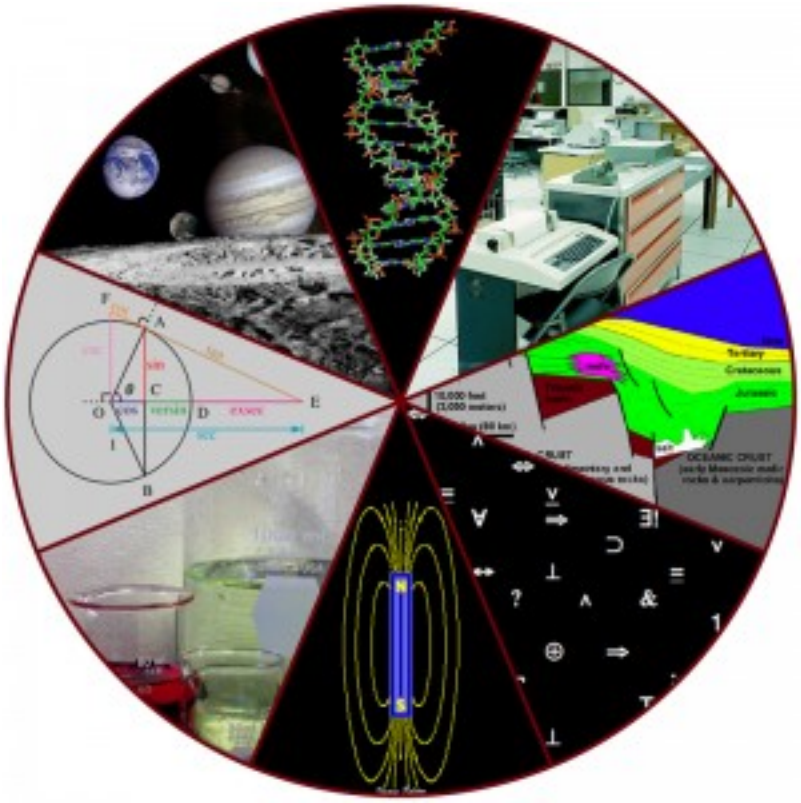


Figure 1.16 Some fields of science include astronomy, biology, computer science, geology, logic, physics, chemistry, and mathematics.

There is no complete agreement when it comes to defining what the natural sciences include. For some experts, the natural sciences are astronomy, biology, chemistry, earth science, and physics. Other scholars choose to divide natural sciences into life sciences, which study living things and include biology, and physical sciences, which study nonliving matter and include astronomy, physics, and chemistry. Some disciplines such as biophysics and biochemistry build on two sciences and are interdisciplinary.

Scientific Inquiry

One thing is common to all forms of science: an ultimate goal “to know.” Curiosity and inquiry are the driving forces for the development of science. Scientists seek to understand the world and the way it operates. Two methods of logical thinking are used: inductive reasoning and deductive reasoning.

Inductive reasoning is a form of logical thinking that uses related observations to arrive at a general conclusion. This type of reasoning is common in descriptive science. A life scientist such as a biologist makes observations and records them. These data can be qualitative (descriptive) or quantitative (consisting of numbers), and the raw data can be supplemented with drawings, pictures, photos, or videos. From many observations, the scientist can infer conclusions (inductions) based on evidence. Inductive reasoning involves formulating generalizations inferred from careful observation and the analysis of a large amount of data. Brain studies often work this way. Many brains are observed while people are doing a task. The part of the brain that lights up, indicating activity, is then demonstrated to be the part controlling the response to that task.

Deductive reasoning or deduction is the type of logic used in hypothesis-based science. In deductive reasoning, the pattern of thinking moves in the opposite direction as compared to inductive reasoning. Deductive reasoning is a form of logical thinking that uses a general principle or law to forecast specific results. From those general principles, a scientist can extrapolate and predict the specific results that would be valid as long as the general principles are valid. For example, a prediction would be that if the climate is becoming warmer in a region, the distribution of plants and animals should change. Comparisons have been made between distributions in the past and the present, and the many changes that have been found are consistent with a warming climate. Finding the change in distribution is evidence that the climate change conclusion is a valid one.

Both types of logical thinking are related to the two main pathways of scientific study: descriptive science and hypothesis-based science. Descriptive (or discovery) science aims to observe, explore, and discover, while hypothesis-based science begins with a specific question or problem and a potential answer or solution that can be tested. The boundary between these two forms of study is often blurred, because most scientific endeavors combine both approaches. Observations lead to questions, questions lead to forming a hypothesis as a possible answer to those questions, and then the hypothesis is tested. Thus, descriptive science and hypothesis-based science are in continuous dialogue.

Hypothesis Testing

Biologists study the living world by posing questions about it and seeking science-based responses. This approach is common to other sciences as well and is often referred to as the scientific method. The scientific method was used even in ancient times, but it was first documented by England’s Sir Francis Bacon (1561–1626), who set up inductive methods for scientific inquiry. The scientific method is not exclusively used by biologists but can be applied to almost anything as a logical problem-solving method.

Figure1.17 Sir Francis Bacon is credited with being the first to document the scientific method.

The scientific process typically starts with an observation (often a problem to be solved) that leads to a question. Let's think about a simple problem that starts with an observation and apply the scientific method to solve the problem. One Monday morning, a student arrives at class and quickly discovers that the classroom is too warm. That is an observation that also describes a problem: the classroom is too warm. The student then asks a question: "Why is the classroom so warm?"

Recall that a hypothesis is a suggested explanation that can be tested. To solve a problem, several hypotheses may be proposed. For example, one hypothesis might be, "The classroom is warm because no one turned on the air conditioning." But there could be other responses to the question, and therefore other hypotheses may be proposed. A second hypothesis might be, "The classroom is warm because there is a power failure, and so the air conditioning doesn't work."

Once a hypothesis has been selected, a prediction may be made. A prediction is similar to a hypothesis but it typically has the format "If . . . then . . . ." For example, the prediction for the first hypothesis might be, "*If* the student turns on the air conditioning, *then* the classroom will no longer be too warm."

A hypothesis must be testable to ensure that it is valid. For example, a hypothesis that depends on what a bear thinks is not testable, because it can never be known what a bear thinks. It should also be falsifiable, meaning that it can be disproven by experimental results. An example of an unfalsifiable hypothesis is "Botticelli's *Birth of Venus* is beautiful." There is no experiment that might show this statement to be false. To test a hypothesis, a researcher will conduct one or more experiments designed to eliminate one or more of the hypotheses. This is important. A hypothesis can be disproven, or

eliminated, but it can never be proven. Science does not deal in proofs like mathematics. If an experiment fails to disprove a hypothesis, then we find support for that explanation, but this is not to say that down the road a better explanation will not be found, or a more carefully designed experiment will be found to falsify the hypothesis.

Each experiment will have one or more variables and one or more controls. A variable is any part of the experiment that can vary or change during the experiment. A control is a part of the experiment that does not change. Look for the variables and controls in the example that follows. As a simple example, an experiment might be conducted to test the hypothesis that phosphate limits the growth of algae in freshwater ponds. A series of artificial ponds are filled with water and half of them are treated by adding phosphate each week, while the other half are treated by adding a salt that is known not to be used by algae. The variable here is the phosphate (or lack of phosphate), the experimental or treatment cases are the ponds with added phosphate and the control ponds are those with something inert added, such as the salt. Just adding something is also a control against the possibility that adding extra matter to the pond has an effect. If the treated ponds show lesser growth of algae, then we have found support for our hypothesis. If they do not, then we reject our hypothesis. Be aware that rejecting one hypothesis does not determine whether or not the other hypotheses can be accepted; it simply eliminates one hypothesis that is not valid . Using the scientific method, the hypotheses that are inconsistent with experimental data are rejected.

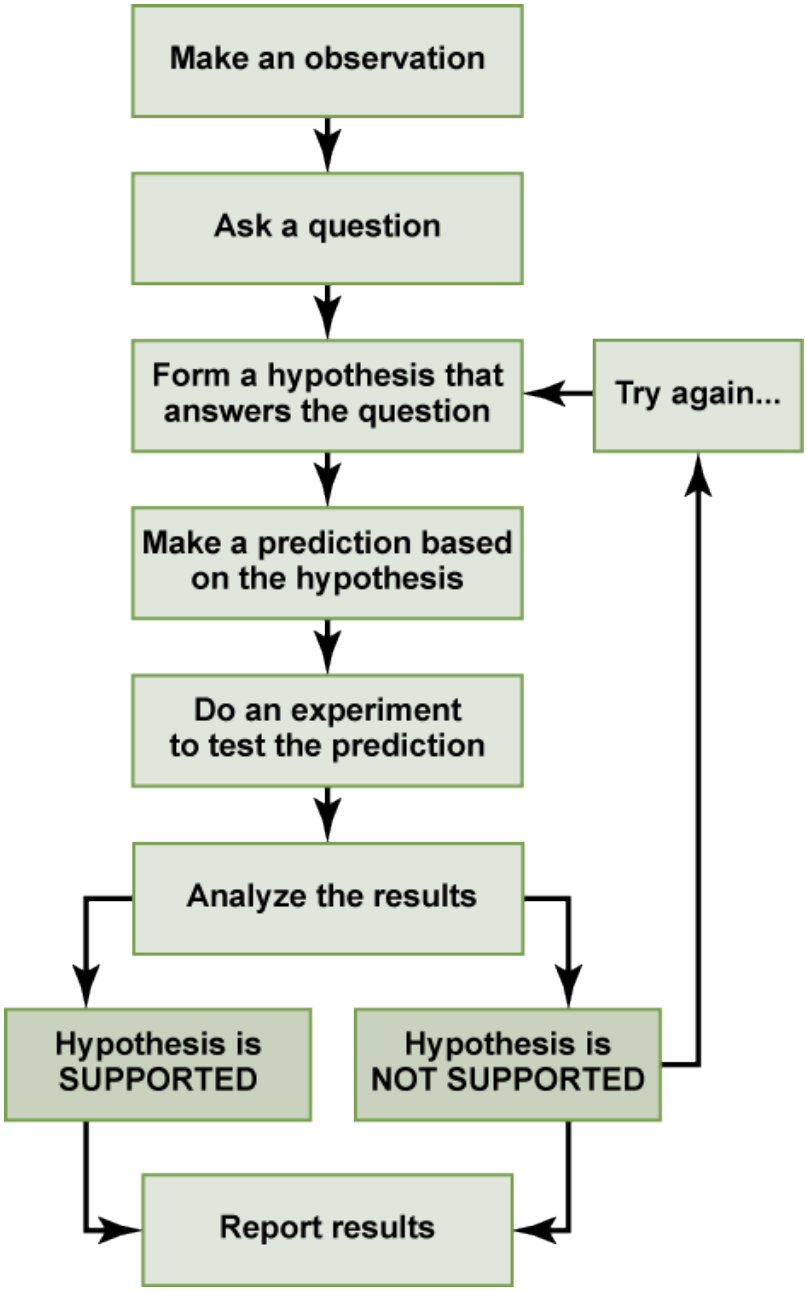


Figure 1.18 The scientific method is a series of defined steps that include experiments and careful observation. If a hypothesis is not supported by data, a new hypothesis can be proposed.

In the example below, the scientific method is used to solve an everyday problem. Which part in the example below is the hypothesis? Which is the prediction? Based on the results of the experiment, is the hypothesis supported? If it is not supported, propose some alternative hypotheses.

1. My toaster doesn't toast my bread.
2. Why doesn't my toaster work?
3. There is something wrong with the electrical outlet.
4. If something is wrong with the outlet, my coffeemaker also won't work when plugged into it.

5. I plug my coffeemaker into the outlet.
6. My coffeemaker works.

In practice, the scientific method is not as rigid and structured as it might at first appear. Sometimes an experiment leads to conclusions that favour a change in approach; often, an experiment brings entirely new scientific questions to the puzzle. Many times, science does not operate in a linear fashion; instead, scientists continually draw inferences and make generalizations, finding patterns as their research proceeds. Scientific reasoning is more complex than the scientific method alone suggests.

Watch a video about the progress of science.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4332#h5p-4

Basic and Applied Science

The scientific community has been debating for the last few decades about the value of different types of science. Is it valuable to pursue science for the sake of simply gaining knowledge, or does scientific knowledge only have worth if we can apply it to solving a specific problem or bettering our lives? This question focuses on the differences between two types of science: basic science and applied science.

Basic science or “pure” science seeks to expand knowledge regardless of the short-term application of that knowledge. It is not focused on developing a product or a service of immediate public or commercial value. The immediate goal of basic science is knowledge for knowledge’s sake, though this does not mean that in the end it may not result in an application.

In contrast, applied science or “technology,” aims to use science to solve real-world problems, making it possible, for example, to improve a crop yield, find a cure for a particular disease, or save animals threatened by a natural disaster. In applied science, the problem is usually defined for the researcher.

Some individuals may perceive applied science as “useful” and basic science as “useless.” A question these people might pose to a scientist advocating knowledge acquisition would be, “What for?” A careful look at the history of science, however, reveals that basic knowledge has resulted in many remarkable applications of great value. Many scientists think that a basic understanding of science is necessary before an application is developed; therefore, applied science relies on the results generated through basic science. Other scientists think that it is time to move on from basic science and instead to find solutions to actual problems. Both approaches are valid. It is true that there are problems that demand immediate attention; however, few solutions would be found without the help of the knowledge generated through basic science.

One example of how basic and applied science can work together to solve practical problems occurred after the discovery of DNA structure led to an understanding of the molecular mechanisms governing

DNA replication. Strands of DNA, unique in every human, are found in our cells, where they provide the instructions necessary for life. During DNA replication, new copies of DNA are made, shortly before a cell divides to form new cells. Understanding the mechanisms of DNA replication enabled scientists to develop laboratory techniques that are now used to identify genetic diseases, pinpoint individuals who were at a crime scene, and determine paternity. Without basic science, it is unlikely that applied science would exist.

Another example of the link between basic and applied research is the Human Genome Project, a study in which each human chromosome was analyzed and mapped to determine the precise sequence of DNA subunits and the exact location of each gene. (The gene is the basic unit of heredity; an individual's complete collection of genes is his or her genome.) Other organisms have also been studied as part of this project to gain a better understanding of human chromosomes. The Human Genome Project relied on basic research carried out with non-human organisms and, later, with the human genome. An important end goal eventually became using the data for applied research seeking cures for genetically related diseases.

Figure 1.19 The Human Genome Project was a 13-year collaborative effort among researchers working in several different fields of science. The project was completed in 2003.

While research efforts in both basic science and applied science are usually carefully planned, it is important to note that some discoveries are made by serendipity, that is, by means of a fortunate accident or a lucky surprise. Penicillin was discovered when biologist Alexander Fleming accidentally left a petri dish of *Staphylococcus* bacteria open. An unwanted mold grew, killing the bacteria. The mold turned out to be *Penicillium*, and a new antibiotic was discovered. Even in the highly organized

world of science, luck—when combined with an observant, curious mind—can lead to unexpected breakthroughs.

Reporting Scientific Work

Whether scientific research is basic science or applied science, scientists must share their findings for other researchers to expand and build upon their discoveries. Communication and collaboration within and between sub disciplines of science are key to the advancement of knowledge in science. For this reason, an important aspect of a scientist's work is disseminating results and communicating with peers. Scientists can share results by presenting them at a scientific meeting or conference, but this approach can reach only the limited few who are present. Instead, most scientists present their results in peer-reviewed articles that are published in scientific journals. Peer-reviewed articles are scientific papers that are reviewed, usually anonymously by a scientist's colleagues, or peers. These colleagues are qualified individuals, often experts in the same research area, who judge whether or not the scientist's work is suitable for publication. The process of peer review helps to ensure that the research described in a scientific paper or grant proposal is original, significant, logical, and thorough. Grant proposals, which are requests for research funding, are also subject to peer review. Scientists publish their work so other scientists can reproduce their experiments under similar or different conditions to expand on the findings. The experimental results must be consistent with the findings of other scientists.

There are many journals and the popular press that do not use a peer-review system. A large number of online open-access journals, journals with articles available without cost, are now available many of which use rigorous peer-review systems, but some of which do not. Results of any studies published in these forums without peer review are not reliable and should not form the basis for other scientific work. In one exception, journals may allow a researcher to cite a personal communication from another researcher about unpublished results with the cited author's permission.

Section Summary

Biology is the science that studies living organisms and their interactions with one another and their environments. Science attempts to describe and understand the nature of the universe in whole or in part. Science has many fields; those fields related to the physical world and its phenomena are considered natural sciences.

A hypothesis is a tentative explanation for an observation. A scientific theory is a well-tested and consistently verified explanation for a set of observations or phenomena. A scientific law is a description, often in the form of a mathematical formula, of the behaviour of an aspect of nature under certain circumstances. Two types of logical reasoning are used in science. Inductive reasoning uses results to produce general scientific principles. Deductive reasoning is a form of logical thinking that predicts results by applying general principles. The common thread throughout scientific research is the use of the scientific method. Scientists present their results in peer-reviewed scientific papers published in scientific journals.

Science can be basic or applied. The main goal of basic science is to expand knowledge without any

expectation of short-term practical application of that knowledge. The primary goal of applied research, however, is to solve practical problems.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4332#h5p-8

Glossary

applied science: a form of science that solves real-world problems

basic science: science that seeks to expand knowledge regardless of the short-term application of that knowledge

control: a part of an experiment that does not change during the experiment

deductive reasoning: a form of logical thinking that uses a general statement to forecast specific results

descriptive science: a form of science that aims to observe, explore, and find things out

falsifiable: able to be disproven by experimental results

hypothesis: a suggested explanation for an event, which can be tested

hypothesis-based science: a form of science that begins with a specific explanation that is then tested

inductive reasoning: a form of logical thinking that uses related observations to arrive at a general conclusion

life science: a field of science, such as biology, that studies living things

natural science: a field of science that studies the physical world, its phenomena, and processes

peer-reviewed article: a scientific report that is reviewed by a scientist's colleagues before publication

physical science: a field of science, such as astronomy, physics, and chemistry, that studies nonliving matter

science: knowledge that covers general truths or the operation of general laws, especially when acquired and tested by the scientific method

scientific law: a description, often in the form of a mathematical formula, for the behavior of some aspect of nature under certain specific conditions

scientific method: a method of research with defined steps that include experiments and careful observation

scientific theory: a thoroughly tested and confirmed explanation for observations or phenomena

variable: a part of an experiment that can vary or change

Media Attribution

- Figure 1.14
    - Cyanobacteria by NASA © Public Domain
    - Stromatolites at Lake Thetis by Ruth Ellison © CC BY-NC (Attribution NonCommercial)
- Figure 1.15 by Eric Erbe, digital colorization by Christopher Pooley, both of USDA, ARS, EMU. © Public Domain
- Figure 1.16 by Image Editor © CC BY (Attribution)
- Figure 1.17 by Paul van Somer I © Public Domain
- Figure 1.19 by U.S. Department of Energy, Human Genome Project © Public Domain

Chapter 2: Introduction to the Chemistry of Life

Figure 2.1 Foods such as bread, fruit, and cheese are rich sources of biological macromolecules.

The elements **carbon, hydrogen, nitrogen, oxygen, sulfur, and phosphorus** are the key building blocks of the chemicals found in living things. They form the **carbohydrates, nucleic acids, proteins, and lipids** (all of which will be defined later in this chapter) that are the fundamental molecular components of all organisms. In this chapter, we will discuss these important building blocks and learn how the unique properties of the atoms of different elements affect their interactions with other atoms to form the molecules of life. These interactions determine what atoms combine and the ultimate shape of the molecules and macromolecules, that shape will determine their function.

Food provides an organism with nutrients—the matter it needs to survive. Many of these critical nutrients come in the form of biological macromolecules, or large molecules necessary for life. These macromolecules are built from different combinations of smaller organic molecules. What specific types of biological macromolecules do living things require? How are these molecules formed? What functions do they serve? In this chapter, we will explore these questions.

Search for Key Points in Chapter 2

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4335#h5p-9

Media Attribution

- Figure 2.1 by Bengt Nyman © CC BY (Attribution)

2.1 The Building Blocks of Molecules

Learning Objectives

By the end of this section, you will be able to:

- Describe matter and elements
- Describe the interrelationship between protons, neutrons, and electrons, and the ways in which electrons can be donated or shared between atoms

Watch a video about electrons and how the electrons in chemical bonds influence the shape and function of molecules.

An interactive H5P element has been excluded from this version of the text. You can view it online here: *https://opentextbc.ca/biology/?p=4343#h5p-10*

At its most fundamental level, life is made up of **matter**. Matter occupies space and has mass. All matter is composed of **elements**, substances that cannot be broken down or transformed chemically into other substances. Each element is made of atoms, each with a constant number of protons and unique properties. A total of 118 elements have been defined; however, only 92 occur naturally, and fewer than 30 are found in living cells. The remaining 26 elements are unstable and, therefore, do not exist for very long or are theoretical and have yet to be detected.

Each element is designated by its chemical symbol (such as H, N, O, C, and Na), and possesses unique properties. These unique properties allow elements to combine and to bond with each other in specific ways.

Atoms

An atom is the smallest component of an element that retains all of the chemical properties of that element. For example, one hydrogen atom has all of the properties of the element hydrogen, such as it exists as a gas at room temperature, and it bonds with oxygen to create a water molecule. Hydrogen atoms cannot be broken down into anything smaller while still retaining the properties of hydrogen. If a

hydrogen atom were broken down into subatomic particles, it would no longer have the properties of hydrogen.

At the most basic level, all organisms are made of a combination of elements. They contain atoms that combine together to form molecules. In multicellular organisms, such as animals, molecules can interact to form cells that combine to form tissues, which make up organs. These combinations continue until entire multicellular organisms are formed.

All atoms contain **protons**, **electrons**, and **neutrons**. The only exception is hydrogen (H), which is made of one proton and one electron. A proton is a positively charged particle that resides in the **nucleus** (the core of the atom) of an atom and has a mass of 1 and a charge of +1. An electron is a negatively charged particle that travels in the space around the nucleus. In other words, it resides outside of the nucleus. It has a negligible mass and has a charge of –1.

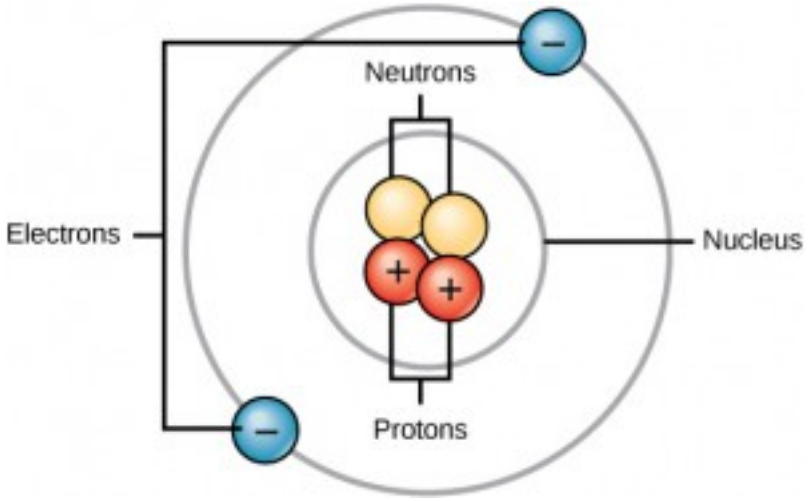


Figure 2.2 Atoms are made up of protons and neutrons located within the nucleus, and electrons surrounding the nucleus.

Neutrons, like protons, reside in the nucleus of an atom. They have a mass of 1 and no charge. The positive (protons) and negative (electrons) charges balance each other in a neutral atom, which has a net zero charge.

Because protons and neutrons each have a mass of 1, the mass of an atom is equal to the number of protons and neutrons of that atom. The number of electrons does not factor into the overall mass, because their mass is so small.

As stated earlier, each element has its own unique properties. Each contains a different number of protons and neutrons, giving it its own atomic number and mass number. The **atomic number** of an element is equal to the number of protons that element contains. The **mass number**, or atomic mass, is the number of protons plus the number of neutrons of that element. Therefore, it is possible to determine the number of neutrons by subtracting the atomic number from the mass number.

These numbers provide information about the elements and how they will react when combined. Different elements have different melting and boiling points, and are in different states (liquid, solid, or gas) at room temperature. They also combine in different ways. Some form specific types of bonds, whereas others do not. How they combine is based on the number of electrons present. Because of

these characteristics, the elements are arranged into the **periodic table of elements**, a chart of the elements that includes the atomic number and relative atomic mass of each element. The periodic table also provides key information about the properties of elements —often indicated by color-coding. The arrangement of the table also shows how the electrons in each element are organized and provides important details about how atoms will react with each other to form molecules.

Isotopes are different forms of the same 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, the most common isotope of carbon, contains six protons and six neutrons. Therefore, it has a mass number of 12 (six protons and six neutrons) and an atomic number of 6 (which makes it carbon). Carbon-14 contains six protons and eight neutrons. Therefore, it has a mass number of 14 (six protons and eight neutrons) and an atomic number of 6, meaning it is still the element carbon. These two alternate forms of carbon are isotopes. Some isotopes are unstable and will lose protons, other subatomic particles, or energy to form more stable elements. These are called **radioactive isotopes** or radioisotopes.

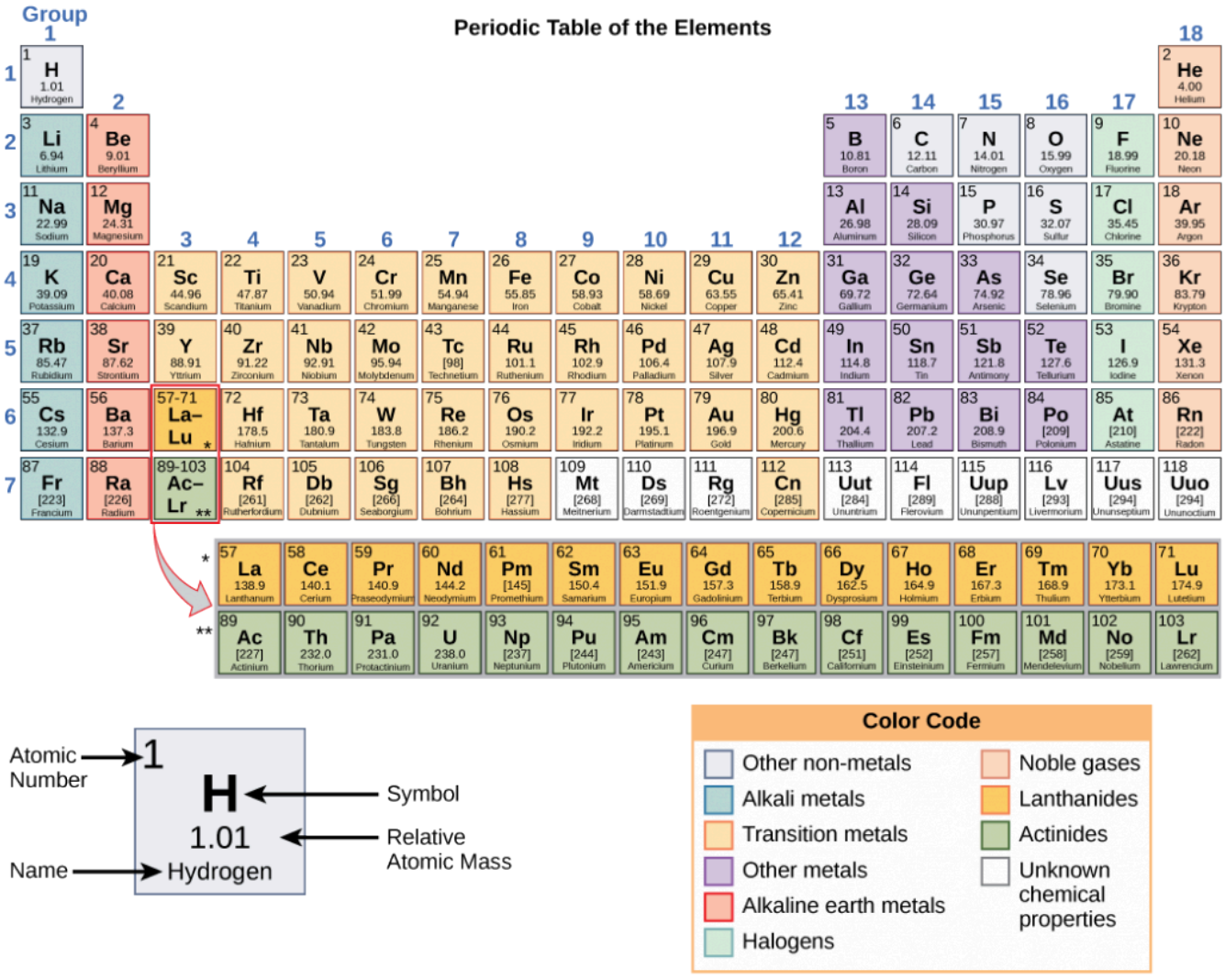


Figure 2.3 Arranged in columns and rows based on the characteristics of the elements, the periodic table provides key information about the elements and how they might interact with each other to form molecules. Most periodic tables provide a key or legend to the information they contain.

How many neutrons do (K) potassium-39 and potassium-40 have, respectively?

Evolution in Action

Carbon Dating

Carbon-14 (^{14}C) is a naturally occurring radioisotope that is created in the atmosphere by cosmic rays. This is a continuous process, so more ^{14}C is always being created. As a living organism develops, the relative level 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 will decline. ^{14}C decays to ^{14}N by a process called beta decay; it gives off energy in this slow process.

After approximately 5,730 years, only one-half of the starting concentration of ^{14}C will have been converted to ^{14}N. The time it takes for half of the original concentration of an isotope to decay to its more stable form is called its half-life. Because the half-life of ^{14}C is long, it is used to age formerly living objects, such as fossils. Using 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. Based on this amount, the age of the fossil can be calculated to about 50,000 years. Isotopes with longer half-lives, such as potassium-40, are used to calculate the ages of older fossils. Through the use of carbon dating, scientists can reconstruct the ecology and biogeography of organisms living within the past 50,000 years.

Figure 2.4 The age of remains that contain carbon and are less than about 50,000 years old, such as this pygmy mammoth, can be determined using carbon dating.

Concept in Action

To learn more about atoms and isotopes, and how you can tell one isotope from another, visit this site and run the simulation.

Chemical Bonds

How elements interact with one another depends on how their electrons are arranged and how many openings for electrons exist at the outermost region where electrons are present in an atom. Electrons exist at energy levels that form shells around the nucleus. The closest shell can hold up to two electrons. The closest shell to the nucleus is always filled first, before any other shell can be filled. Hydrogen has one electron; therefore, it has only one spot occupied within the lowest shell. Helium has two electrons; therefore, it can completely fill the lowest shell with its two electrons. If you look at the periodic table, you will see that hydrogen and helium are the only two elements in the first row. This is because they only have electrons in their first shell. Hydrogen and helium are the only two elements that have the lowest shell and no other shells.

The second and third energy levels can hold up to eight electrons. The eight electrons are arranged in four pairs and one position in each pair is filled with an electron before any pairs are completed.

Looking at the periodic table again, you will notice that there are seven rows. These rows correspond to the number of shells that the elements within that row have. The elements within a particular row have increasing numbers of electrons as the columns proceed from left to right. Although each element has the same number of shells, not all of the shells are completely filled with electrons. If you look at the second row of the periodic table, you will find lithium (Li), beryllium (Be), boron (B), carbon (C), nitrogen (N), oxygen (O), fluorine (F), and neon (Ne). These all have electrons that occupy only the first and second shells. Lithium has only one electron in its outermost shell, beryllium has two electrons, boron has three, and so on, until the entire shell is filled with eight electrons, as is the case with neon.

Not all elements have enough electrons to fill their outermost shells, but an atom is at its most stable when all of the electron positions in the outermost shell are filled. Because of these vacancies in the outermost shells, we see the formation of **chemical bonds**, or interactions between two or more of the same or different elements that result in the formation of molecules. To achieve greater stability, atoms will tend to completely fill their outer shells and will bond with other elements to accomplish this goal by sharing electrons, accepting electrons from another atom, or donating electrons to another atom. Because the outermost shells of the elements with low atomic numbers (up to calcium, with atomic

number 20) can hold eight electrons, this is referred to as the **octet rule**. An element can donate, accept, or share electrons with other elements to fill its outer shell and satisfy the octet rule.

When an atom does not contain equal numbers of protons and electrons, it is called an **ion**. Because the number of electrons does not equal the number of protons, each ion has a net charge. Positive ions are formed by losing electrons and are called **cations**. Negative ions are formed by gaining electrons and are called **anions**.

For example, sodium only has one electron in its outermost 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 and only 10 electrons, leaving it with an overall charge of +1. It is now called a sodium ion.

The chlorine atom 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 and 18 electrons, giving it a net negative (–1) charge. It is now called a chloride ion. This movement of electrons from one element to another is referred to as **electron transfer**. As illustrates, a sodium atom (Na) only has one electron in its outermost shell, whereas a chlorine atom (Cl) has seven electrons in its outermost shell. A sodium atom will donate its one electron to empty its shell, and a chlorine atom will accept that electron to fill its shell, becoming chloride. 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) or –1 (chloride) charge.

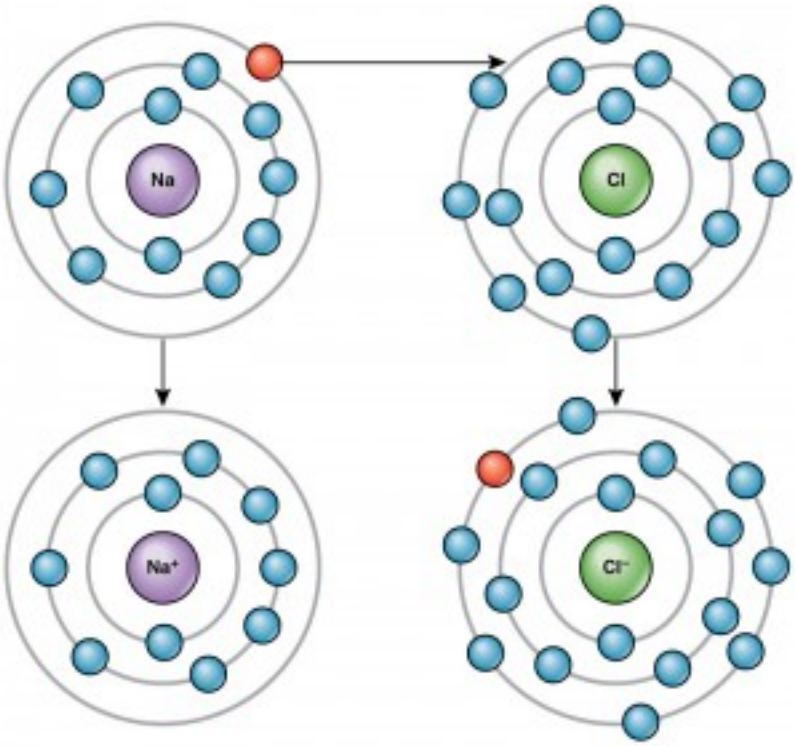


Figure 2.5 Elements tend to fill their outermost shells with electrons. To do this, they can either donate or accept electrons from other elements.

Ionic Bonds

There are four types of bonds or interactions: ionic, covalent, hydrogen bonds, and van der Waals interactions. Ionic and covalent bonds are strong interactions that require a larger energy input to break apart. When an element donates an electron from its outer shell, as in the sodium atom example above, a positive ion is formed. The element accepting the electron is now negatively charged. Because positive and negative charges attract, these ions stay together and form an **ionic bond**, or a bond between ions. The elements bond together with the electron from one element staying predominantly with the other element. When Na^+ and Cl^- ions combine to produce NaCl, an electron from a sodium atom stays with the other seven from the chlorine atom, and the sodium and chloride ions attract each other in a lattice of ions with a net zero charge.

Covalent Bonds

Another type of strong chemical bond between two or more atoms is a **covalent bond**. These bonds form when a pair of electrons is shared between two elements and are the strongest and most common form of chemical bond in living organisms. Covalent bonds form between the elements that make up the biological molecules in our cells. Unlike ionic bonds, covalent bonds do not dissociate in water.

The hydrogen and oxygen atoms that combine to form water molecules are bound together by covalent bonds. The electron from the hydrogen atom divides its time between the outer shell of the hydrogen atom and the incomplete outer shell of the oxygen atom. To completely fill the outer shell of an oxygen atom, two electrons from two hydrogen atoms are needed, hence the subscript “2” in H_2O. The electrons are shared between the atoms, dividing their time between them to “fill” the outer shell of each. This sharing is a lower energy state for all of the atoms involved than if they existed without their outer shells filled.

There are two types of covalent bonds: polar and nonpolar. **Nonpolar covalent bonds** form between two atoms of the same element or between different elements that share the electrons equally. For example, an oxygen atom can bond with another oxygen atom to fill their outer shells. This association is nonpolar because the electrons will be equally distributed between each oxygen atom. Two covalent bonds form between the two oxygen atoms because oxygen requires two shared electrons to fill its outermost shell. Nitrogen atoms will form three covalent bonds (also called triple covalent) between two atoms of nitrogen because each nitrogen atom needs three electrons to fill its outermost shell. Another example of a nonpolar covalent bond is found in the methane (CH_4) molecule. The carbon atom 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. These elements all share the electrons equally, creating four nonpolar covalent bonds.

In a **polar covalent bond**, the electrons shared by the atoms spend more time closer to one nucleus than to the other nucleus. Because of the unequal distribution of electrons between the different nuclei, a slightly positive ($\delta+$) or slightly negative ($\delta-$) charge develops. The covalent bonds between hydrogen and oxygen atoms in water are polar covalent bonds. The shared electrons spend more time near the oxygen nucleus, giving it a small negative charge, than they spend near the hydrogen nuclei, giving these molecules a small positive charge.

Hydrogen Bonds

Ionic and covalent bonds are strong bonds that require considerable energy to break. However, not all bonds between elements are ionic or covalent bonds. Weaker bonds can also form. These are attractions that occur between positive and negative charges that do not require much energy to break. Two weak bonds that occur frequently are **hydrogen bonds** and van der Waals interactions. These bonds give rise to the unique properties of water and the unique structures of DNA and proteins.

When polar covalent bonds containing a hydrogen atom form, the hydrogen atom in that bond has a slightly positive charge. This is because the shared electron is pulled more strongly toward the other element and away from the hydrogen nucleus. Because the hydrogen atom is slightly positive (δ+), it will be attracted to neighboring negative partial charges (δ–). When this happens, a weak interaction occurs between the δ+ charge of the hydrogen atom of one molecule and the δ– charge of the other molecule. This interaction is called a hydrogen bond. This type of bond is common; for example, the liquid nature of water is caused by the hydrogen bonds between water molecules. Hydrogen bonds give water the unique properties that sustain life. If it were not for hydrogen bonding, water would be a gas rather than a liquid at room temperature.

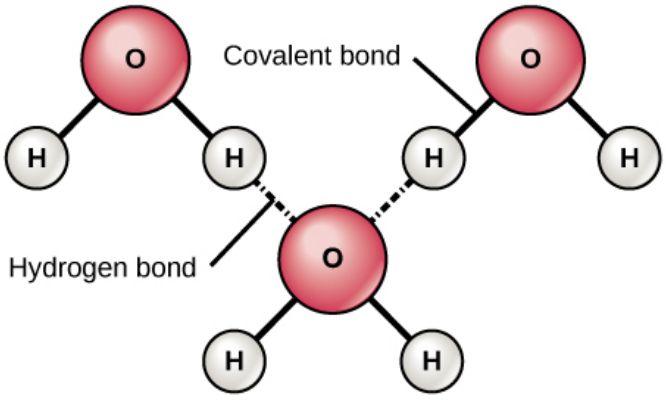


Figure 2.6 Hydrogen bonds form between slightly positive (δ+) and slightly negative (δ–) charges of polar covalent molecules, such as water.

Hydrogen bonds can form between different molecules and they do not always have to include a water molecule. Hydrogen atoms in polar bonds within any molecule can form bonds with other adjacent molecules. For example, hydrogen bonds hold together two long strands of DNA to give the DNA molecule its characteristic double-stranded structure. Hydrogen bonds are also responsible for some of the three-dimensional structure of proteins.

van der Waals Interactions

Like hydrogen bonds, **van der Waals interactions** are weak attractions or interactions between molecules. They occur between polar, covalently bound, atoms in different molecules. Some of these weak attractions are caused by temporary partial charges formed when electrons move around a nucleus. These weak interactions between molecules are important in biological systems.

Radiography Technicians

Have you or anyone you know ever had a magnetic resonance imaging (MRI) scan, a mammogram, or an X-ray? These tests produce images of your soft tissues and organs (as with an MRI or mammogram) or your bones (as happens in an X-ray) by using either radio waves or special isotopes (radiolabeled or fluorescently labeled) that are ingested or injected into the body. These tests provide data for disease diagnoses by creating images of your organs or skeletal system.

MRI imaging works by subjecting hydrogen nuclei, which are abundant in the water in soft tissues, to fluctuating magnetic fields, which cause them to emit their own magnetic field. This signal is then read by sensors in the machine and interpreted by a computer to form a detailed image.

Some radiography technologists and technicians specialize in computed tomography, MRI, and mammography. They produce films or images of the body that help medical professionals examine and diagnose. Radiologists work directly with patients, explaining machinery, preparing them for exams, and ensuring that their body or body parts are positioned correctly to produce the needed images. Physicians or radiologists then analyze the test results.

Radiography technicians can work in hospitals, doctors' offices, or specialized imaging centers. Training to become a radiography technician happens at hospitals, colleges, and universities that offer certificates, associate's degrees, or bachelor's degrees in radiography.

Section Summary

Matter is anything that occupies space and has mass. It is made up of atoms of different elements. All of the 92 elements that occur naturally have unique qualities that allow them to combine in various ways to create compounds or molecules. Atoms, which consist of protons, neutrons, and electrons, are the smallest units of an element that retain all of the properties of that element. Electrons can be donated or shared between atoms to create bonds, including ionic, covalent, and hydrogen bonds, as well as van der Waals interactions.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4343#h5p-11

Glossary

anion: a negative ion formed by gaining electrons

atomic number: the number of protons in an atom

cation: a positive ion formed by losing electrons

chemical bond: an interaction between two or more of the same or different elements that results in the formation of molecules

covalent bond: a type of strong bond between two or more of the same or different elements; forms when electrons are shared between elements

electron: a negatively charged particle that resides outside of the nucleus in the electron orbital; lacks functional mass and has a charge of –1

electron transfer: the movement of electrons from one element to another

element: one of 118 unique substances that cannot be broken down into smaller substances and retain the characteristic of that substance; each element has a specified number of protons and unique properties

hydrogen bond: a weak bond between partially positively charged hydrogen atoms and partially negatively charged elements or molecules

ion: an atom or compound that does not contain equal numbers of protons and electrons, and therefore has a net charge

ionic bond: a chemical bond that forms between ions of opposite charges

isotope: one or more forms of an element that have different numbers of neutrons

mass number: the number of protons plus neutrons in an atom

matter: anything that has mass and occupies space

neutron: a particle with no charge that resides in the nucleus of an atom; has a mass of 1

nonpolar covalent bond: a type of covalent bond that forms between atoms when electrons are shared equally between atoms, resulting in no regions with partial charges as in polar covalent bonds

nucleus: (chemistry) the dense center of an atom made up of protons and (except in the case of a hydrogen atom) neutrons

octet rule: states that the outermost shell of an element with a low atomic number can hold eight electrons

periodic table of elements: an organizational chart of elements, indicating the atomic number and mass number of each element; also provides key information about the properties of elements

polar covalent bond:a type of covalent bond in which electrons are pulled toward one atom and away from another, resulting in slightly positive and slightly negative charged regions of the molecule

proton: a positively charged particle that resides in the nucleus of an atom; has a mass of 1 and a charge of +1

radioactive isotope: an isotope that spontaneously emits particles or energy to form a more stable element

van der Waals interaction: a weak attraction or interaction between molecules caused by slightly positively charged or slightly negatively charged atoms

Media Attribution

- Figure 2.4 by Bill Faulkner/NPS

2.2 Water

Learning Objectives

By the end of this section, you will be able to:

- Describe the properties of water that are critical to maintaining life

Watch a video about why we need oxygen and how it causes problems for living things.

An interactive H5P element has been excluded from this version of the text. You can view it online here: *https://opentextbc.ca/biology/?p=4351#h5p-12*

Do you ever wonder why scientists spend time looking for water on other planets? It is because water is essential to life; even minute traces of it on another planet can indicate that life could or did exist on that planet. Water is one of the more abundant molecules in living cells and the one most critical to life as we know it. Approximately 60–70 percent of your body is made up of water. Without it, life simply would not exist.

Water Is Polar

The hydrogen and oxygen atoms within water molecules form polar covalent bonds. The shared electrons spend more time associated with the oxygen atom than they do with hydrogen atoms. There is no overall charge to a water molecule, but there is a slight positive charge on each hydrogen atom and a slight negative charge on the oxygen atom. Because of these charges, the slightly positive hydrogen atoms repel each other and form the unique shape. Each water molecule attracts other water molecules because of the positive and negative charges in the different parts of the molecule. Water also attracts other polar molecules (such as sugars), forming hydrogen bonds. When a substance readily forms hydrogen bonds with water, it can dissolve in water and is referred to as **hydrophilic** (“water-loving”). Hydrogen bonds are not readily formed with nonpolar substances like oils and fats . These nonpolar compounds are **hydrophobic** (“water-fearing”) and will not dissolve in water.

Figure 2.7 As this macroscopic image of oil and water shows, oil is a nonpolar compound and, hence, will not dissolve in water. Oil and water do not mix.

Water Stabilizes Temperature

The hydrogen bonds in water allow it to absorb and release heat energy more slowly than many other substances. **Temperature** is a measure of the motion (kinetic energy) of molecules. As the motion increases, energy is higher and thus temperature is higher. Water absorbs a great deal of energy before its temperature rises. Increased energy disrupts the hydrogen bonds between water molecules. Because these bonds can be created and disrupted rapidly, water absorbs an increase in energy and temperature changes only minimally. This means that water moderates temperature changes within organisms and in their environments. As energy input continues, the balance between hydrogen-bond formation and destruction swings toward the destruction side. More bonds are broken than are formed. This process results in the release of individual water molecules at the surface of the liquid (such as a body of water, the leaves of a plant, or the skin of an organism) in a process called **evaporation**. Evaporation of sweat, which is 90 percent water, allows for cooling of an organism, because breaking hydrogen bonds requires an input of energy and takes heat away from the body.

Conversely, as molecular motion decreases and temperatures drop, less energy is present to break the hydrogen bonds between water molecules. These bonds remain intact and begin to form a rigid, lattice-like structure (e.g., ice) (Figure 2.8 **a**). When frozen, ice is less dense than liquid water (the molecules are farther apart). This means that ice floats on the surface of a body of water (Figure 2.8 **b**). In lakes, ponds, and oceans, ice will form on the surface of the water, creating an insulating barrier to protect the animal and plant life beneath from freezing in the water. If this did not happen, plants and animals living in water would freeze in a block of ice and could not move freely, making life in cold temperatures difficult or impossible.

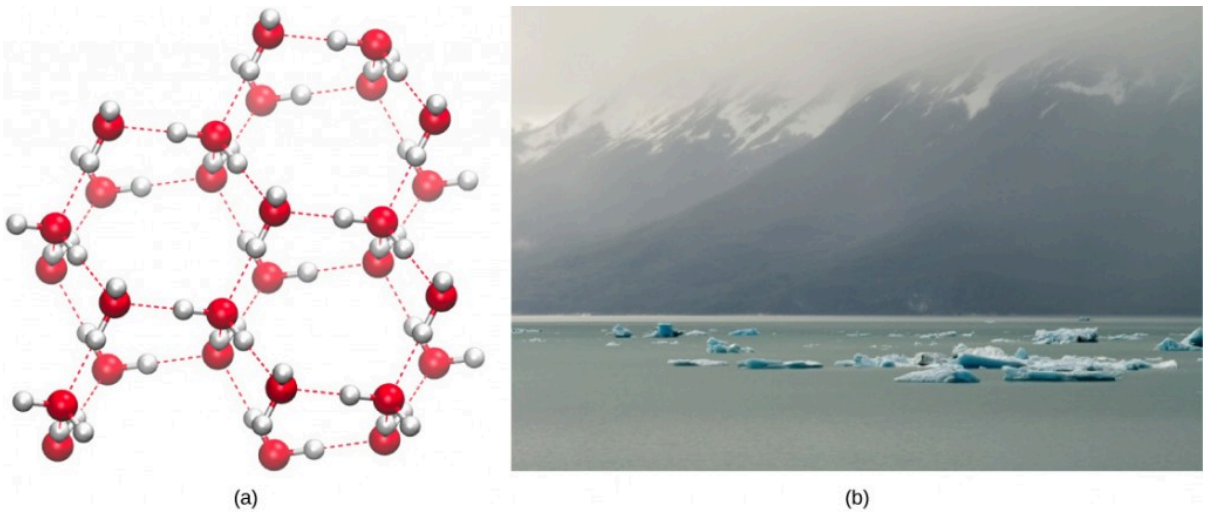


Figure 2.8 (a) The lattice structure of ice makes it less dense than the freely flowing molecules of liquid water. Ice's lower density enables it to (b) float on water. (credit a: modification of work by Jane Whitney; credit b: modification of work by Carlos Ponte)

Water Is an Excellent Solvent

Because water is polar, with slight positive and negative charges, ionic compounds and polar molecules can readily dissolve in it. Water is, therefore, what is referred to as a **solvent**—a substance capable of dissolving another substance. The charged particles will form hydrogen bonds with a surrounding layer of water molecules. This is referred to as a sphere of hydration and serves to keep the particles separated or dispersed in the water. In the case of table salt (NaCl) mixed in water, the sodium and chloride ions separate, or dissociate, in the water, and spheres of hydration are formed around the ions. A positively charged sodium ion is surrounded by the partially negative charges of oxygen atoms in water molecules. A negatively charged chloride ion is surrounded by the partially positive charges of hydrogen atoms in water molecules. These spheres of hydration are also referred to as hydration shells. The polarity of the water molecule makes it an effective solvent and is important in its many roles in living systems.

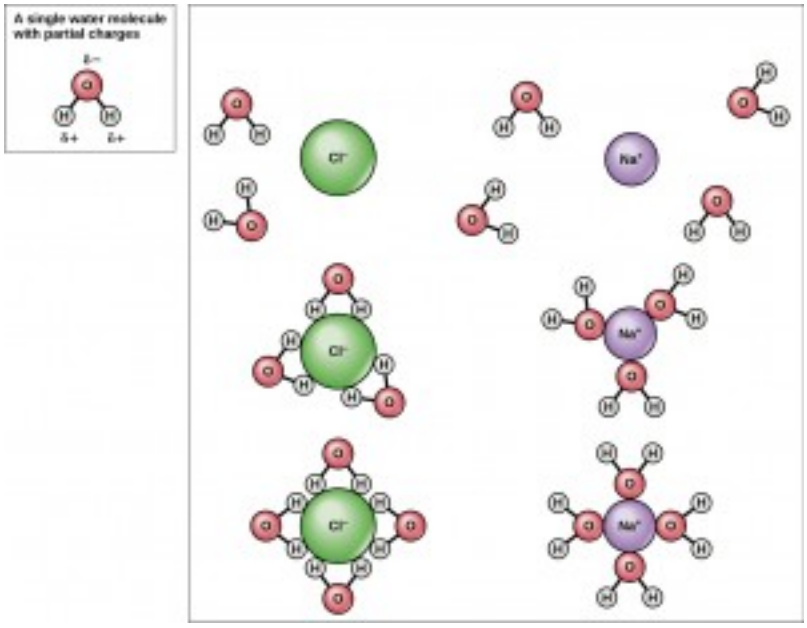


Figure 2.9 When table salt (NaCl) is mixed in water, spheres of hydration form around the ions.

Water Is Cohesive

Have you ever filled up a glass of water to the very top and then slowly added a few more drops? Before it overflows, the water actually 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-air (gas) interface, although there is no more room in the glass. Cohesion gives rise to **surface tension**, the capacity of a substance to withstand rupture when placed under tension or stress. When you drop a small scrap of paper onto a droplet of water, the paper floats on top of the water droplet, although the object is denser (heavier) than the water. This occurs because of the surface tension that is created by the water molecules. Cohesion and surface tension keep the water molecules intact and the item floating on the top. It is even possible to “float” a steel needle on top of a glass of water if you place it gently, without breaking the surface tension.

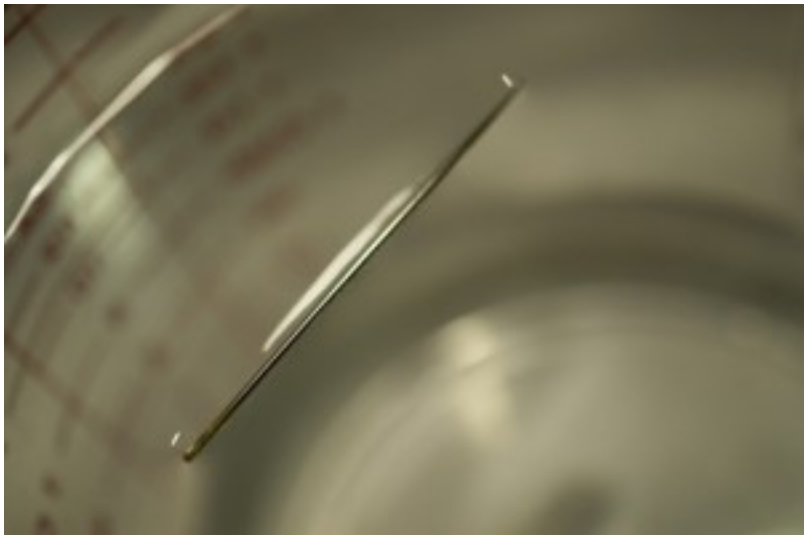


Figure 2.10 The weight of a needle on top of water pulls the surface tension downward; at the same time, the surface tension of the water is pulling it up, suspending the needle on the surface of the water and keeping it from sinking. Notice the indentation in the water around the needle.

These cohesive forces are also related to the water's property of **adhesion**, or the attraction between water molecules and other molecules. This is observed when water "climbs" up a straw placed in a glass of water. You will notice that the water appears to be higher on the sides of the straw than in the middle. This is because the water molecules are attracted to the straw and therefore adhere to it.

Cohesive and adhesive forces are important for sustaining life. For example, because of these forces, water can flow up from the roots to the tops of plants to feed the plant.

Concept in Action

To learn more about water, visit the U.S. Geological Survey Water Science for Schools: All About Water! website.

Buffers, pH, Acids, and Bases

The pH of a solution is a measure of its acidity or alkalinity. You have probably used **litmus paper**, 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 or base (alkalinity) exists in a solution. You might have even used some to make sure

the water in an outdoor swimming pool is properly treated. In both cases, this pH test measures the amount of hydrogen ions that exists in a given solution. High concentrations of hydrogen ions yield a low pH, whereas low levels of hydrogen ions result in a high pH. The overall concentration of hydrogen ions is inversely related to its pH and can be measured on the **pH scale** (Figure 2.11). Therefore, the more hydrogen ions present, the lower the pH; conversely, the fewer hydrogen ions, the higher the pH.

The pH scale ranges from 0 to 14. A change of one unit on the pH scale represents a change in the concentration of hydrogen ions by a factor of 10, a change in two units represents a change in the concentration of hydrogen ions by a factor of 100. Thus, small changes in pH represent large changes in the concentrations of hydrogen ions. Pure water is neutral. It is neither acidic nor basic, and has a pH of 7.0. 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. The blood in your veins is slightly alkaline (pH = 7.4). The environment in your stomach is highly acidic (pH = 1 to 2). Orange juice is mildly acidic (pH = approximately 3.5), whereas baking soda is basic (pH = 9.0).

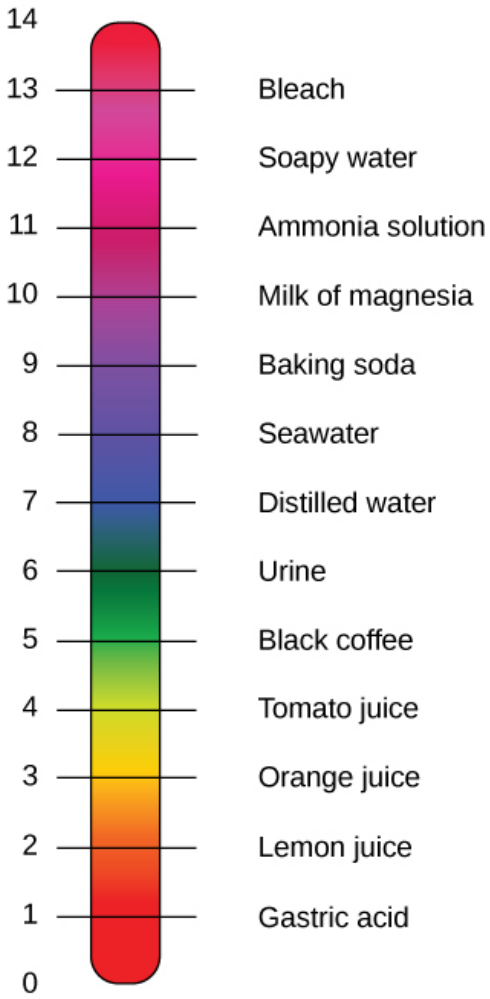


Figure 2.11 The pH scale measures the amount of hydrogen ions (H+) in a substance.

Acids are substances that provide hydrogen ions (H^+) and lower pH, whereas **bases** provide hydroxide

ions (OH^-) and raise pH. The stronger the acid, the more readily it donates H^+. For example, hydrochloric acid and lemon juice are very acidic and readily give up H^+ when added to water. Conversely, bases are those substances that readily donate OH^-. The OH^- ions combine with H^+ to produce water, which raises a substance's pH. Sodium hydroxide and many household cleaners are very alkaline and give up OH^- rapidly when placed in water, thereby raising the pH.

Most cells in our bodies operate within a very narrow window of the pH scale, typically ranging only from 7.2 to 7.6. If the pH of the body is outside of this range, the respiratory system malfunctions, as do other organs in the body. Cells no longer function properly, and proteins will break down. Deviation outside of the pH range can induce coma or even cause death.

So how is it that we can ingest or inhale acidic or basic substances and not die? Buffers are the key. **Buffers** readily absorb excess H^+ or OH^-, keeping the pH of the body carefully maintained in the aforementioned narrow range. Carbon dioxide is part of a prominent buffer system in the human body; it keeps the pH within the proper range. This buffer system involves carbonic acid (H_2CO_3) and bicarbonate (HCO_3^-) anion. If too much H^+ enters the body, bicarbonate will combine with the H^+ to create carbonic acid and limit the decrease in pH. Likewise, if too much OH^- is introduced into the system, carbonic acid will rapidly dissociate into bicarbonate and H^+ ions. The H^+ ions can combine with the OH^- ions, limiting the increase in pH. While carbonic acid is an important product in this reaction, its presence is fleeting because the carbonic acid is released from the body as carbon dioxide gas each time we breathe. Without this buffer system, the pH in our bodies would fluctuate too much and we would fail to survive.

Section Summary

Water has many properties that are critical to maintaining life. It is polar, allowing for the formation of hydrogen bonds, which allow ions and other polar molecules to dissolve in water. Therefore, water is an excellent solvent. The hydrogen bonds between water molecules give water the ability to hold heat better than many other substances. As the temperature rises, the hydrogen bonds between water continually break and reform, allowing for the overall temperature to remain stable, although increased energy is added to the system. Water's cohesive forces allow for the property of surface tension. All of these unique properties of water are important in the chemistry of living organisms.

The pH of a solution is a measure of the concentration of hydrogen ions in the solution. A solution with a high number of hydrogen ions is acidic and has a low pH value. A solution with a high number of hydroxide ions is basic and has a high pH value. The pH scale ranges from 0 to 14, with a pH of 7 being neutral. Buffers are solutions that moderate pH changes when an acid or base is added to the buffer system. Buffers are important in biological systems because of their ability to maintain constant pH conditions.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here: https://opentextbc.ca/biology/?p=4351#h5p-13

Glossary

acid: a substance that donates hydrogen ions and therefore lowers pH

adhesion: the attraction between water molecules and molecules of a different substance

base: a substance that absorbs hydrogen ions and therefore raises pH

buffer: a solution that resists a change in pH by absorbing or releasing hydrogen or hydroxide ions

cohesion: the intermolecular forces between water molecules caused by the polar nature of water; creates surface tension

evaporation: the release of water molecules from liquid water to form water vapor

hydrophilic: describes a substance that dissolves in water; water-loving

hydrophobic: describes a substance that does not dissolve in water; water-fearing

litmus paper: filter paper that has been treated with a natural water-soluble dye so it can be used as a pH indicator

pH scale: a scale ranging from 0 to 14 that measures the approximate concentration of hydrogen ions of a substance

solvent: a substance capable of dissolving another substance

surface tension: the cohesive force at the surface of a body of liquid that prevents the molecules from separating

temperature: a measure of molecular motion

References

Humphrey, W., Dalke, A. and Schulten, K., “VMD—Visual Molecular Dynamics”, *J. Molec. Graphics*, 1996, vol. 14, pp. 33-38. http://www.ks.uiuc.edu/Research/vmd/

Media Attribution

- Figure 2.7 by Gautam Dogra
- Figure 2.8
    - ice lattice by Jane Whitney
    - (b) by Carlos Ponte
- Figure 2.10 by Cory Zanker
- Figure 2.11 by Edward Stevens

2.3 Biological Molecules

Learning Objectives

By the end of this section, you will be able to:

- Describe the ways in which carbon is critical to life
- Explain the impact of slight changes in amino acids on organisms
- Describe the four major types of biological molecules
- Understand the functions of the four major types of molecules

Watch a video about proteins and protein enzymes.

An interactive H5P element has been excluded from this version of the text. You can view it online here: *https://opentextbc.ca/biology/?p=4368#h5p-14*

The large molecules necessary for life that are built from smaller organic molecules are called biological **macromolecules**. There are four major classes of biological macromolecules (carbohydrates, lipids, proteins, and nucleic acids), and each is an important component of the cell and performs a wide array of functions. Combined, these molecules make up the majority of a cell's mass. Biological macromolecules are organic, meaning that they contain carbon. In addition, they may contain hydrogen, oxygen, nitrogen, phosphorus, sulfur, and additional minor elements.

Carbon

It is often said that life is "carbon-based." This means that carbon atoms, bonded to other carbon atoms or other elements, form the fundamental components of many, if not most, of the molecules found uniquely in living things. Other elements play important roles in biological molecules, but carbon certainly qualifies as the "foundation" element for molecules in living things. It is the bonding properties of carbon atoms that are responsible for its important role.

Carbon Bonding

Carbon contains four electrons in its outer shell. Therefore, it can form four covalent bonds with other atoms or molecules. The simplest organic carbon molecule is methane (CH_4), in which four hydrogen atoms bind to a carbon atom.

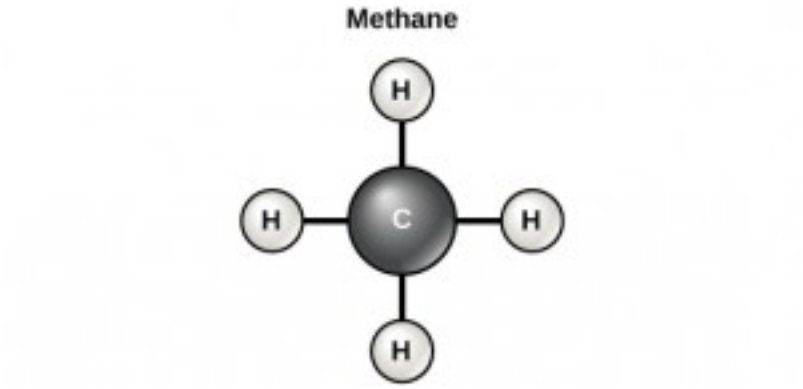


Figure 2.12 Carbon can form four covalent bonds to create an organic molecule. The simplest carbon molecule is methane (CH4), depicted here.

However, structures that are more complex are made using carbon. Any of the hydrogen atoms can be replaced with another carbon atom covalently bonded to the first carbon atom. In this way, long and branching chains of carbon compounds can be made (Figure 2.13 **a**). The carbon atoms may bond with atoms of other elements, such as nitrogen, oxygen, and phosphorus (Figure 2.13 **b**). The molecules may also form rings, which themselves can link with other rings (Figure 2.13 **c**). This diversity of molecular forms accounts for the diversity of functions of the biological macromolecules and is based to a large degree on the ability of carbon to form multiple bonds with itself and other atoms.

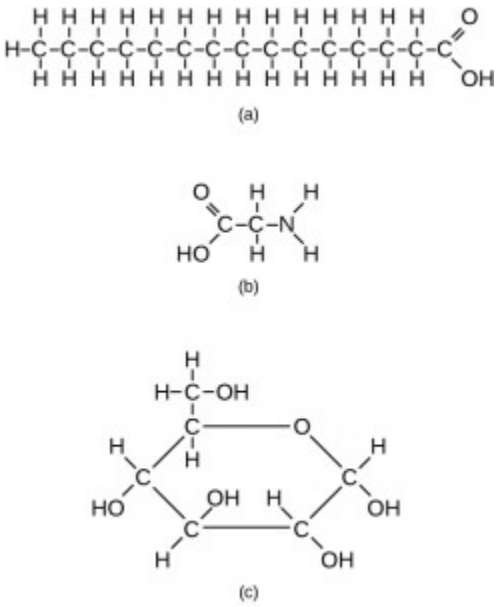


Figure 2.13 These examples show three molecules (found in living organisms) that contain carbon atoms bonded in various ways to other carbon atoms and the atoms of other elements. (a) This molecule of stearic acid has a long chain of carbon atoms. (b) Glycine, a component of proteins, contains carbon, nitrogen, oxygen, and hydrogen atoms. (c) Glucose, a sugar, has a ring of carbon atoms and one oxygen atom.

Carbohydrates

Carbohydrates are macromolecules with which most consumers are somewhat familiar. To lose weight, some individuals adhere to “low-carb” diets. Athletes, in contrast, often “carb-load” before important competitions to ensure that they have sufficient 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. Carbohydrates also have other important functions in humans, animals, and plants.

Carbohydrates can be represented by the formula $(CH_2O)_n$, where n is the number of carbon atoms in the molecule. In other words, the ratio of carbon to hydrogen to oxygen is 1:2:1 in carbohydrate molecules. Carbohydrates are classified into three subtypes: monosaccharides, disaccharides, and polysaccharides.

Monosaccharides (mono- = “one”; sacchar- = “sweet”) are simple sugars, the most common of which is glucose. In monosaccharides, the number of carbon atoms usually ranges from three to six. Most monosaccharide names end with the suffix -ose. Depending on the number of carbon atoms in the sugar, they may be known as trioses (three carbon atoms), pentoses (five carbon atoms), and hexoses (six carbon atoms).

Monosaccharides may exist as a linear chain or as ring-shaped molecules; in aqueous solutions, they are usually found in the ring form.

The chemical formula for glucose is $C_6H_{12}O_6$. In most living species, 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 by the process of photosynthesis, and the glucose, in turn, is used for the energy requirements of the plant. The excess synthesized glucose is often stored as starch that is broken down by other organisms that feed on plants.

Galactose (part of lactose, or milk sugar) and fructose (found 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 differing arrangements of atoms in the carbon chain.

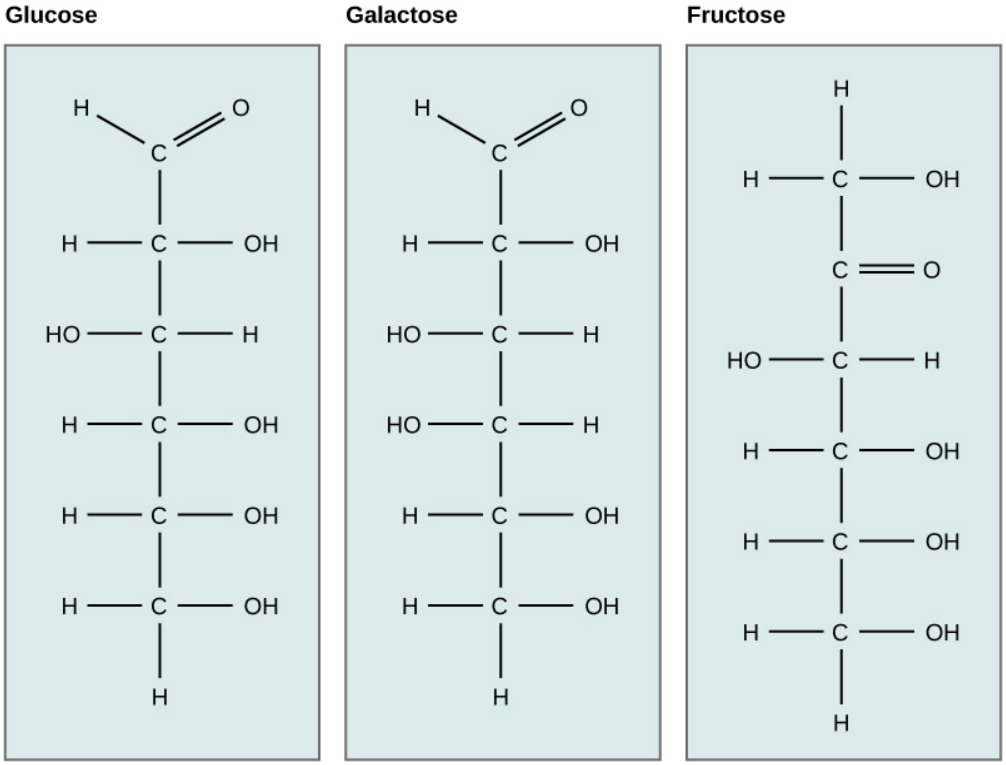


Figure 2.14 Glucose, galactose, and fructose are isomeric monosaccharides, meaning that they have the same chemical formula but slightly different structures.

Disaccharides (di- = “two”) form when two monosaccharides undergo a dehydration reaction (a reaction in which the removal of a water molecule occurs). During this process, the hydroxyl group (–OH) of one monosaccharide combines with a hydrogen atom of another monosaccharide, releasing a molecule of water (H_2O) and forming a covalent bond between atoms in the two sugar molecules.

Common disaccharides include lactose, maltose, and sucrose. 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 from a dehydration reaction between two glucose molecules. The most common disaccharide is sucrose, or table sugar, which is composed of the monomers glucose and fructose.

A long chain of monosaccharides linked by covalent bonds is known as a **polysaccharide** (poly- = “many”). The chain may be branched or unbranched, and it may contain different types of monosaccharides. Polysaccharides may be very large molecules. Starch, glycogen, cellulose, and chitin are examples of polysaccharides.

Starch is the stored form of sugars in plants and is made up of amylose and amylopectin (both polymers of glucose). Plants are able to synthesize glucose, and the excess glucose is stored as starch in different plant parts, including roots and seeds. The starch that is consumed by animals is broken down into smaller molecules, such as glucose. The cells can then absorb the glucose.

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 glucose levels decrease, glycogen is broken down to release glucose.

Cellulose is one of the most abundant natural biopolymers. The cell walls of plants are mostly made of cellulose, which 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 bonds between particular carbon atoms in the glucose molecule.

Every other glucose monomer in cellulose is flipped over and packed tightly as extended long chains. This gives cellulose its rigidity and high tensile strength—which is so important to plant cells. Cellulose passing through our digestive system is called dietary fiber. While the glucose-glucose bonds in cellulose cannot be broken down by human digestive enzymes, herbivores such as cows, buffalos, and horses are able to digest grass that is rich in cellulose and use it as a food source. In these animals, certain species of bacteria reside in the rumen (part of the digestive system of herbivores) and secrete the enzyme cellulase. The appendix also contains bacteria that break down 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.

Carbohydrates serve other functions in different animals. Arthropods, such as insects, spiders, and crabs, have an outer skeleton, called the exoskeleton, which protects their internal body parts. This exoskeleton is made of the biological macromolecule **chitin**, which is a nitrogenous carbohydrate. It is made of repeating units of a modified sugar containing nitrogen.

Thus, through differences in molecular structure, carbohydrates are able to serve the very different functions of energy storage (starch and glycogen) and structural support and protection (cellulose and chitin).

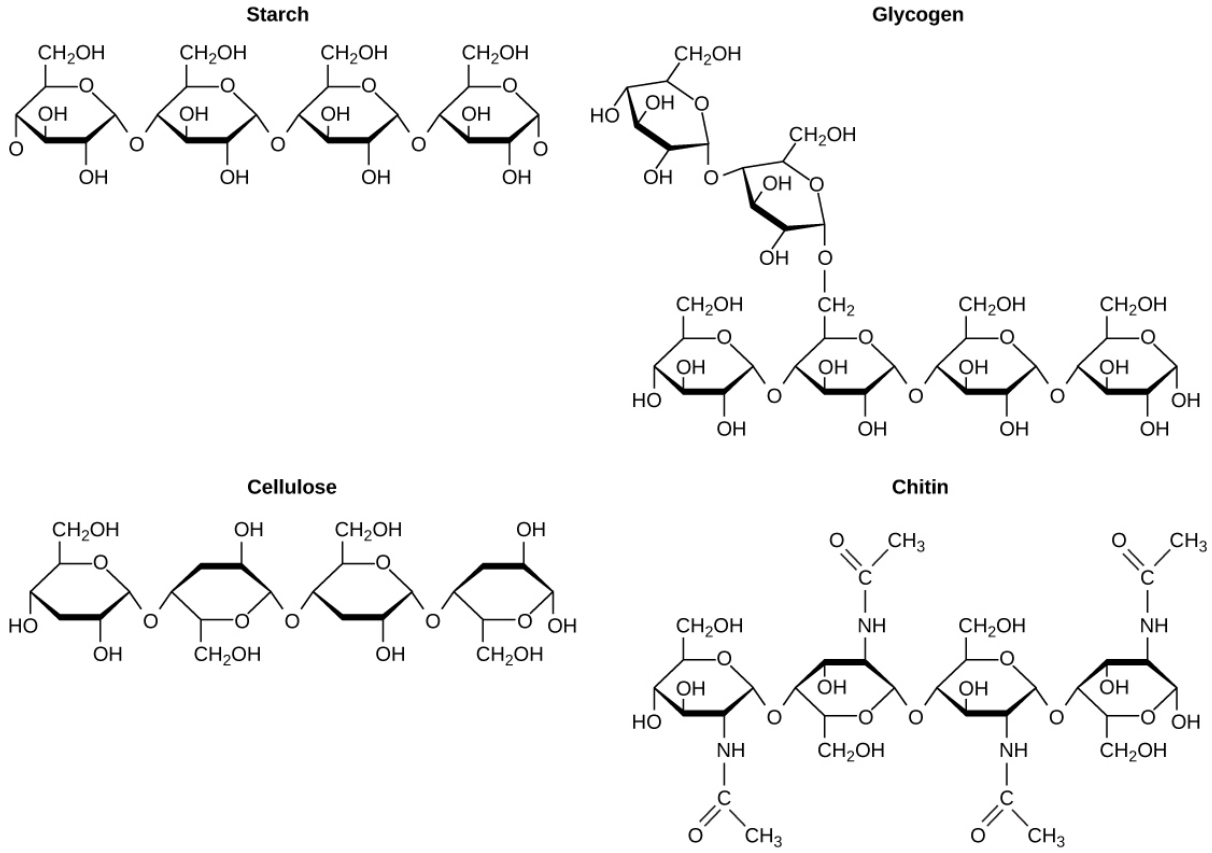


Figure 2.15 Although their structures and functions differ, all polysaccharide carbohydrates are made up of monosaccharides and have the chemical formula (CH2O)n.

Registered Dietitian: Obesity is a worldwide health concern, and many diseases, such as diabetes and heart disease, are becoming more prevalent because of obesity. This is one of the reasons why registered dietitians are increasingly sought after for advice. Registered dietitians help plan food and nutrition programs for individuals in various settings. They often work with patients in health-care facilities, designing nutrition plans to prevent and treat diseases. For example, dietitians may teach a patient with diabetes how to manage blood-sugar levels by eating the correct types and amounts of carbohydrates. Dietitians may also work in nursing homes, schools, and private practices.

To become a registered dietitian, one needs to earn at least a bachelor’s degree in dietetics, nutrition, food technology, or a related field. In addition, registered dietitians must complete a supervised internship program and pass a national exam. Those who pursue careers in dietetics take courses in nutrition, chemistry, biochemistry, biology, microbiology, and human physiology. Dietitians must become experts in the chemistry and functions of food (proteins, carbohydrates, and fats).

Through the Indigenous Lens (Suzanne Wilkerson and Charles Molnar)

I work at Camosun College located in beautiful Victoria, British Columbia with campuses on the Traditional Territories of the Lekwungen and W̱SÁNEĆ peoples. The underground storage bulb of the camas flower shown below has been an important food source for many of the Indigenous peoples of Vancouver Island and throughout the western area of North America. Camas bulbs are still eaten as a

traditional food source and the preparation of the camas bulbs relates to this text section about carbohydrates.

Figure 2.16 Image of a blue camas flower and an insect pollinator. The underground bulb of camas is baked in a fire pit. Heat acts like pancreatic amylase enzyme and breaks down long chains of indigestible inulin into digestible mono and di-saccharides.

Most often plants create starch as the stored form of carbohydrate. Some plants, like camas create inulin. Inulin is used as dietary fibre however, it is not readily digested by humans. If you were to bite into a raw camas bulb it would taste bitter and has a gummy texture. The method used by Indigenous peoples to make camas both digestible and tasty is to bake the bulbs slowly for a long period in an underground firepit covered with specific leaves and soil. The heat acts like our pancreatic amylase enzyme and breaks down the long chains of inulin into digestible mono and di-saccharides.

Properly baked, the camas bulbs taste like a combination of baked pear and cooked fig. It is important to note that while the blue camas is a food source, it should not be confused with the white death camas, which is particularly toxic and deadly. The flowers look different, but the bulbs look very similar.

Lipids

Lipids include a diverse group of compounds that are united by a common feature. Lipids are hydrophobic (“water-fearing”), or insoluble in water, because they are nonpolar molecules. This is because they are hydrocarbons that include only nonpolar carbon-carbon or carbon-hydrogen bonds. Lipids perform many different functions in a cell. Cells store energy for long-term use in the form of lipids called **fats**. Lipids also provide insulation from the environment for plants and animals. For example, they help keep aquatic birds and mammals dry because of their water-repelling nature. Lipids are also the building blocks of many hormones and are an important constituent of the plasma membrane. Lipids include fats, oils, waxes, phospholipids, and steroids.

Figure 2.17 Hydrophobic lipids in the fur of aquatic mammals, such as this river otter, protect them from the elements.

A fat molecule, such as a triglyceride, consists of two main components—glycerol and fatty acids. Glycerol is an organic compound with three carbon atoms, five hydrogen atoms, and three hydroxyl (–OH) groups. Fatty acids have a long chain of hydrocarbons to which an acidic 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, a fatty acid is attached to each of the three oxygen atoms in the –OH groups of the glycerol molecule with a covalent bond.

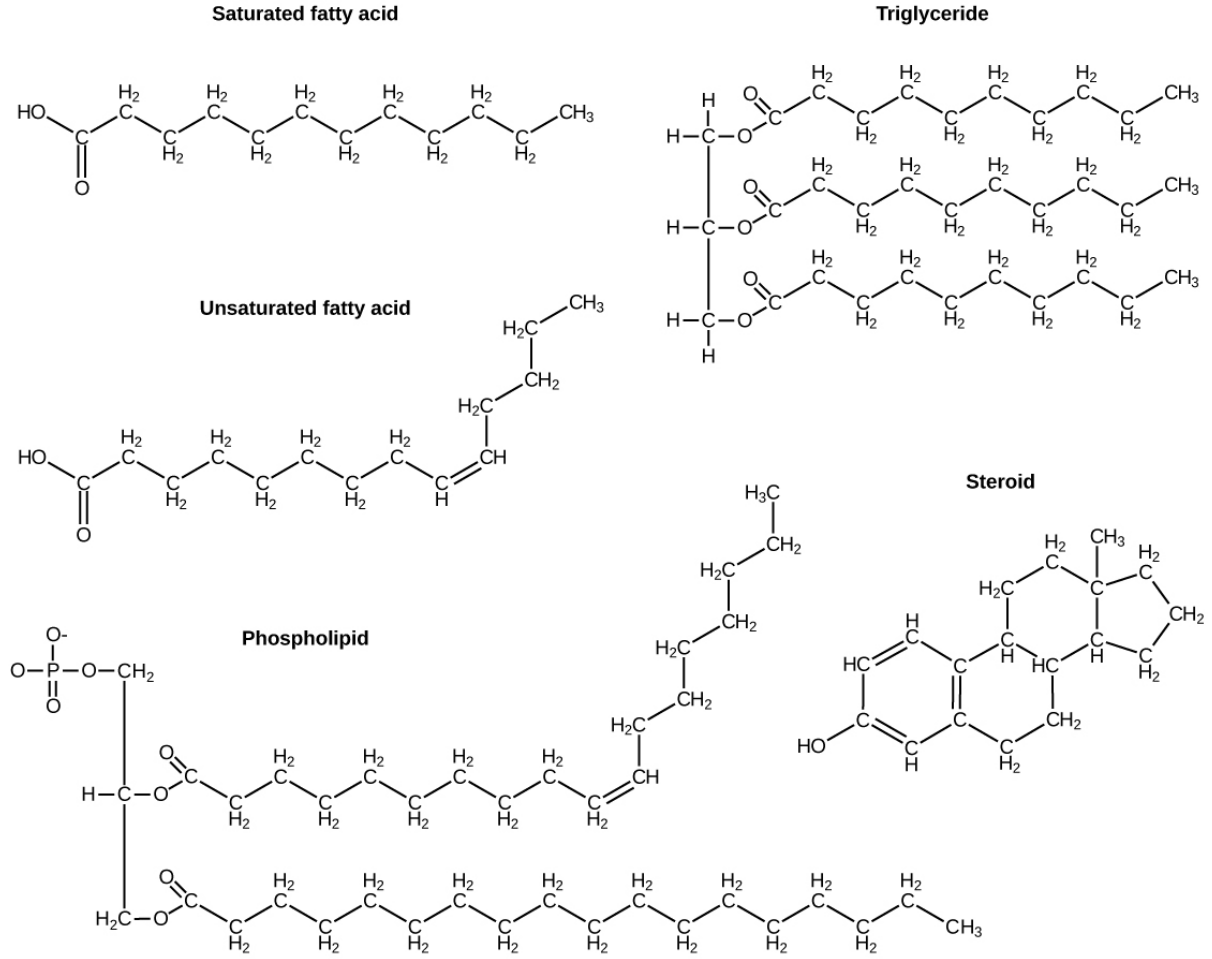


Figure 2.18 Lipids include fats, such as triglycerides, which are made up of fatty acids and glycerol, phospholipids, and steroids.

During this covalent bond formation, three water molecules are released. The three fatty acids in the fat may be similar or dissimilar. These fats are also called **triglycerides** because they have three fatty acids. 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 hypogaea*, the scientific name for 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 saturated. **Saturated fatty acids** are saturated with hydrogen; in other words, the number of hydrogen atoms attached to the carbon skeleton is maximized.

When the hydrocarbon chain contains a double bond, the fatty acid is an **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).

Saturated fats tend to get packed tightly and are solid at room temperature. Animal fats with stearic acid and palmitic acid contained in meat, and the fat with butyric acid contained in butter, are examples

of saturated fats. Mammals store fats in specialized cells called adipocytes, where globules of fat occupy most of the cell. In plants, fat or oil is stored in seeds and is used as a source of energy during embryonic development.

Unsaturated fats or oils are usually of plant origin and contain unsaturated fatty acids. The double bond causes a bend or a "kink" that prevents the fatty acids from packing tightly, keeping them liquid at room temperature. Olive oil, corn oil, canola oil, and cod liver oil are examples of unsaturated fats. Unsaturated fats help to improve blood cholesterol levels, whereas saturated fats contribute to plaque formation in the arteries, which increases the risk of a heart attack.

In the food industry, oils are artificially hydrogenated to make them semi-solid, leading to less spoilage and increased shelf life. 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. This forms a ***trans*-fat** from a *cis*-fat. The orientation of the double bonds affects the chemical properties of the fat.

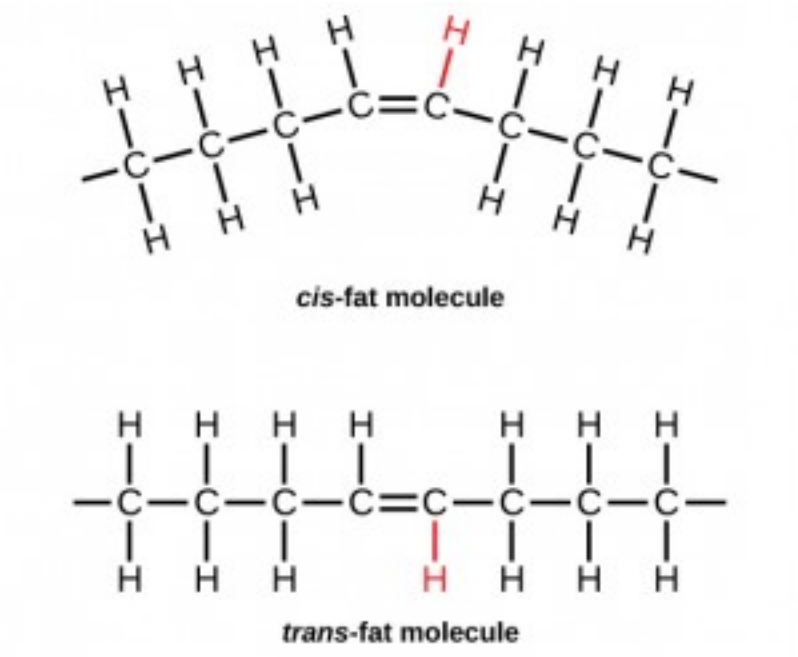


Figure 2.19 During the hydrogenation process, the orientation around the double bonds is changed, making a trans-fat from a cis-fat. This changes the chemical properties of the molecule.

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 lipoprotein (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 eliminated the use of *trans*-fats, and U.S. food labels are now required to list their *trans*-fat content.

Essential fatty acids are fatty acids that are required but not synthesized by the human body. Consequently, they must be supplemented through the diet. **Omega-3 fatty acids** fall into this category and are one of only two known essential fatty acids for humans (the other being omega-6 fatty acids).

They are a type of polyunsaturated fat and are called omega-3 fatty acids because the third carbon from the end of the fatty acid participates in a double bond.

Salmon, trout, and tuna are good sources of omega-3 fatty acids. Omega-3 fatty acids are important in brain function and normal growth and development. They may also prevent heart disease and reduce the risk of cancer.

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. Fats serve as long-term energy storage. They also provide insulation for the body. Therefore, "healthy" unsaturated fats in moderate amounts should be consumed on a regular basis.

Phospholipids are the major constituent of the plasma membrane. Like fats, they are composed of fatty acid chains attached to a glycerol or similar backbone. Instead of three fatty acids attached, however, there are two fatty acids and the third carbon of the glycerol backbone is bound to a phosphate group. The phosphate group is modified by the addition of an alcohol.

A phospholipid has both hydrophobic and hydrophilic regions. The fatty acid chains are hydrophobic and exclude themselves from water, whereas the phosphate is hydrophilic and interacts with water.

Cells are surrounded by a membrane, which has a bilayer of phospholipids. The fatty acids of phospholipids face inside, away from water, whereas the phosphate group can face either the outside environment or the inside of the cell, which are both aqueous.

Through the Indigenous Lens

For the First peoples of the Pacific Northwest the fat rich fish ooligan, with 20% fat by body weight, was a crucial part of the diet of several First Nations. Why? Because fat is the most calorie dense food and having a storable, high calorie compact energy source would be important to survival. The nature of its fat also made it an important trade good. Like salmon, ooligan returns to its birth stream after years at sea. Its arrival in the early spring made it the first fresh food of the year. In the Tsimshianic languages the arrival of the ooligan … was traditionally announced with the cry, 'Hlaa aat'ixshi halimootxw!' … meaning 'Our Saviour has just arrived!'

Figure 2.20 Image of cooked ooligan. With 20% fat by body weight, this fat rich fish is a crucial part of the First Nations diet.

As you learned above all fats are hydrophobic (water hating). To isolate the fat, the fish is boiled and the floating fat skimmed off. Ooligan fat composition is 30% saturated fat (like butter) and 55% monounsaturated fat (like plant oils). Importantly it is a solid grease at room temperature. Because it is low in polyunsaturated fats (which oxidize and spoil quickly) it can be stored for later use and used as a trade item. Its composition is said to make it as healthy as olive oil, or better as it has omega 3 fatty acids that reduce risk for diabetes and stroke. It also is rich in three fat soluble vitamins A, E and K.

Steroids and Waxes

Unlike the phospholipids and fats discussed earlier, **steroids** have a ring structure. Although they do not resemble other lipids, they are grouped with them because they are also hydrophobic. All steroids have four, linked carbon rings and several of them, like cholesterol, have a short tail.

Cholesterol is a steroid. Cholesterol is mainly synthesized in the liver and is the precursor of many steroid hormones, such as testosterone and estradiol. It is also the precursor of vitamins E and K. Cholesterol is the precursor of bile salts, which help in the breakdown of fats and their subsequent absorption by cells. Although cholesterol is often spoken of in negative terms, it is necessary for the proper functioning of the body. It is a key component of the plasma membranes of animal cells.

Waxes are made up of a hydrocarbon chain with an alcohol (–OH) group and a fatty acid. Examples of animal waxes include beeswax and lanolin. Plants also have waxes, such as the coating on their leaves, that helps prevent them from drying out.

Concept in Action

For an additional perspective on lipids, explore “Biomolecules: The Lipids” through this interactive animation.

Proteins

Proteins are one of the most abundant organic molecules in living systems and have the most diverse range of functions of all macromolecules. 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 different 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.

The functions of proteins are very diverse because there are 20 different chemically distinct amino acids that form long chains, and the amino acids can be in any order. For example, proteins can function as enzymes or hormones. **Enzymes**, which are produced by living cells, are catalysts in biochemical reactions (like digestion) and are usually proteins. Each enzyme is specific for the substrate (a reactant that binds to an enzyme) upon which it acts. Enzymes can function to break molecular bonds, to rearrange bonds, or to form new bonds. An example of an enzyme is salivary amylase, which breaks down amylose, a component of starch.

Hormones are chemical signaling molecules, usually proteins or steroids, secreted by an endocrine gland or group of 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 maintains blood glucose levels.

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. Changes in temperature, pH, and exposure to chemicals may lead to permanent changes in the shape of the protein, leading to a loss of function or denaturation (to be discussed in more detail later). All proteins are made up of different arrangements of the same 20 kinds of 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 bonded to an amino group ($–NH_2$), a carboxyl group (–COOH), and a hydrogen atom. Every amino acid also has another variable atom or group of atoms

bonded to the central carbon atom known as the R group. The R group is the only difference in structure between the 20 amino acids; otherwise, the amino acids are identical.

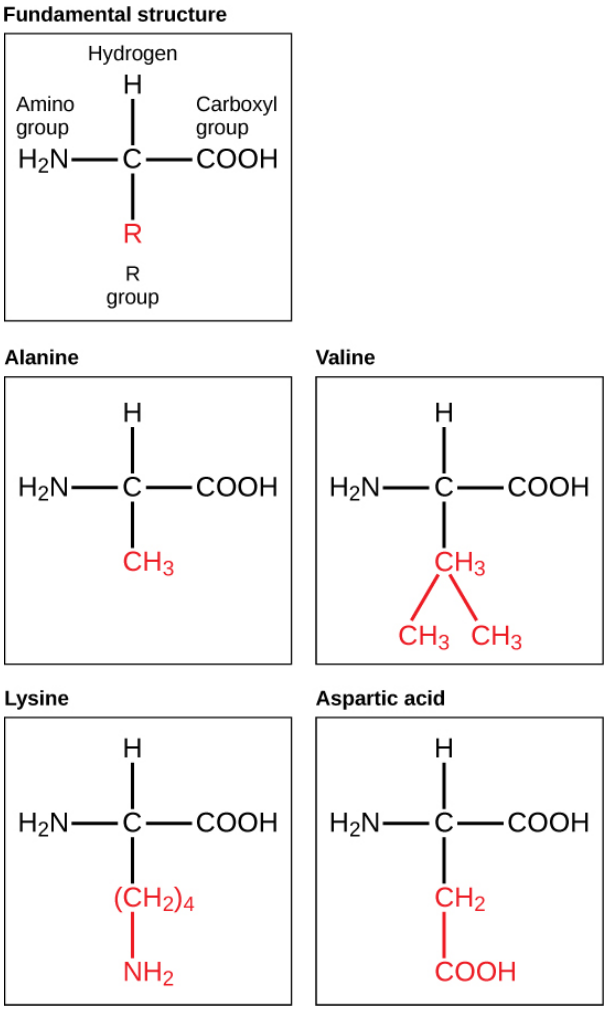


Figure 2.21 Amino acids are made up of a central carbon bonded to an amino group (–NH2), a carboxyl group (–COOH), and a hydrogen atom. The central carbon's fourth bond varies among the different amino acids, as seen in these examples of alanine, valine, lysine, and aspartic acid.

The chemical nature of the R group determines the chemical nature of the amino acid within its protein (that is, whether it is acidic, basic, polar, or nonpolar).

The sequence and number of amino acids ultimately determine a 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 dehydration reaction. The carboxyl group of one amino acid and the amino group of a second amino acid combine, releasing a water molecule. The resulting bond is the peptide bond.

The products formed by such a linkage are called **polypeptides**. 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, have a distinct shape, and have a unique function.

Evolution in Action

The Evolutionary Significance of Cytochrome cCytochrome c is an important component of the molecular machinery that harvests energy from glucose. Because this 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 sequence similarity among cytochrome c molecules of different species; evolutionary relationships can be assessed by measuring the similarities or differences among various species' protein sequences.

For example, scientists have determined that human cytochrome c contains 104 amino acids. For each cytochrome c molecule that has been sequenced to date from different organisms, 37 of these amino acids appear in the same position in each cytochrome c. This indicates that all of these organisms are descended from a common ancestor. On comparing the human and chimpanzee protein sequences, no sequence difference was found. When human and rhesus monkey sequences were compared, a single difference was found in one amino acid. In contrast, human-to-yeast comparisons show a difference in 44 amino acids, suggesting that humans and chimpanzees have a more recent common ancestor than humans and the rhesus monkey, or humans and yeast.

Protein Structure

As discussed earlier, the shape of a protein is critical to its function. 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**.

The unique sequence and number of amino acids in a polypeptide chain is its primary structure. The unique sequence for every protein is ultimately determined by the gene that encodes the protein. Any change in the gene sequence may lead to a different amino acid being added to the polypeptide chain, causing a change in protein structure and function. In sickle cell anemia, the hemoglobin β chain has a single amino acid substitution, causing a change in both the structure and function of the protein. 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—that dramatically decreases life expectancy in the affected individuals—is a single amino acid of the 600.

Because of this change of one amino acid in the chain, the normally biconcave, or disc-shaped, red blood cells assume a crescent or "sickle" shape, which clogs arteries. This can lead to a myriad of serious health problems, such as breathlessness, dizziness, headaches, and abdominal pain for those who have this disease.

Folding patterns resulting from interactions between the non-R group portions of amino acids give rise to the secondary structure of the protein. The most common are the alpha (α)-helix and beta (β)-pleated sheet structures. Both structures are held in shape by hydrogen bonds. In the alpha helix, the bonds form between every fourth amino acid and cause a twist in the amino acid 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 to each other, and hydrogen bonds form between the same pairs of atoms on each of the aligned amino acids. The α-helix and β-pleated sheet structures are found in many globular and fibrous proteins.

The unique three-dimensional structure of a polypeptide is known as its tertiary structure. This structure is caused by chemical interactions between various amino acids and regions of the polypeptide. Primarily, the interactions among R groups create the complex three-dimensional tertiary structure of a protein. There may be ionic bonds formed between R groups on different amino acids, or hydrogen bonding beyond that involved in the secondary structure. 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.

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, hemoglobin is a combination of four polypeptide subunits.

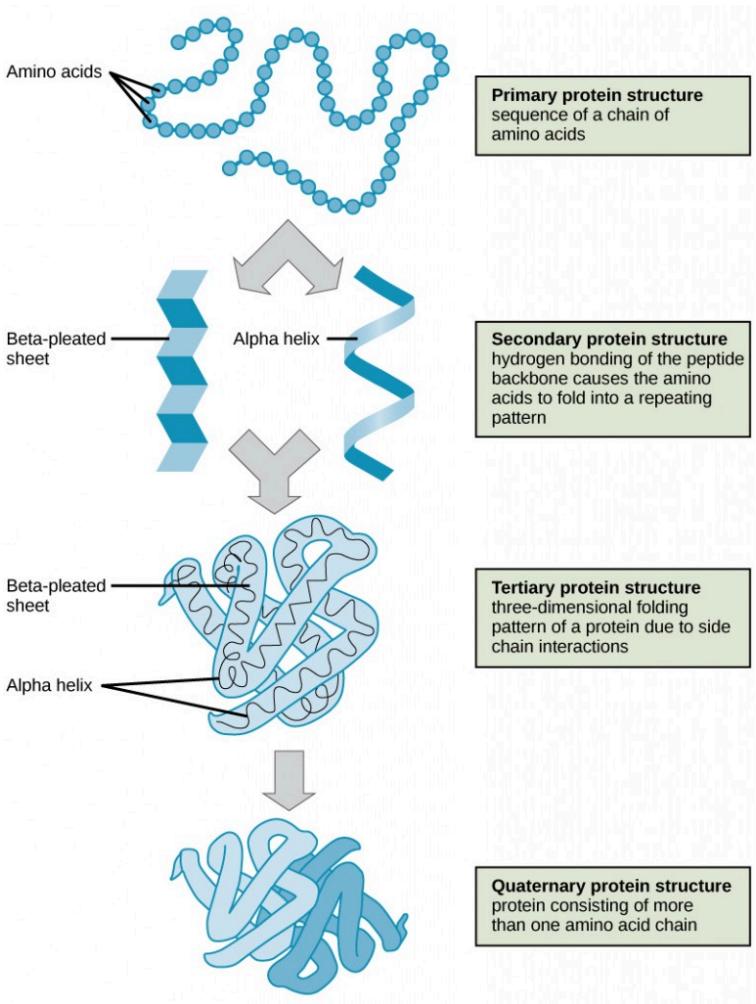


Figure 2.22 The four levels of protein structure can be observed in these illustrations.

Each protein has its own unique sequence and shape held together by chemical interactions. If the protein is subject to changes in temperature, pH, or exposure to chemicals, the protein structure may change, losing its shape in what is known as **denaturation** as discussed earlier. Denaturation is often reversible because the primary structure is preserved if the denaturing agent is removed, allowing the protein to resume its function. Sometimes denaturation is irreversible, leading to a loss of function. One example of protein denaturation can be seen when an egg is fried or boiled. The albumin protein in the liquid egg white is denatured when placed in a hot pan, changing from a clear substance to an opaque white substance. Not all proteins are denatured at high temperatures; for instance, bacteria that survive in hot springs have proteins that are adapted to function at those temperatures.

Concept in Action

For an additional perspective on proteins, explore “Biomolecules: The Proteins” through this interactive animation.

Nucleic Acids

Nucleic acids are key macromolecules in the continuity of life. They carry the genetic blueprint of a cell and carry instructions for the functioning of the cell.

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.

The other type of nucleic acid, RNA, is mostly involved in protein synthesis. The DNA molecules never leave the nucleus, but instead use an RNA intermediary to communicate with the rest of the cell. Other types of RNA are also 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, DNA or RNA. Each nucleotide is made up of three components: a nitrogenous base, a pentose (five-carbon) sugar, and a phosphate group . Each nitrogenous base in a nucleotide is attached to a sugar molecule, which is attached to a phosphate group.

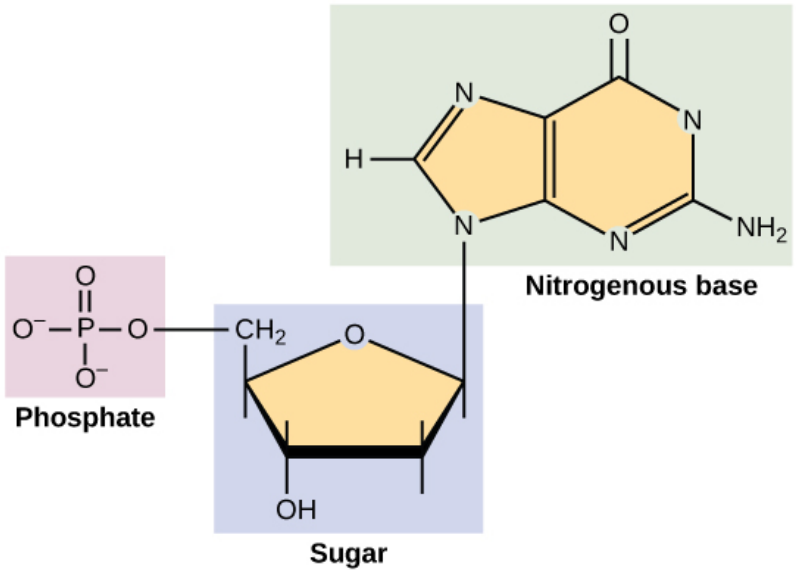


Figure 2.23 A nucleotide is made up of three components: a nitrogenous base, a pentose sugar, and a phosphate group.

DNA Double-Helical Structure

DNA has a double-helical structure. It is composed of two strands, or polymers, of nucleotides. The strands are formed with bonds between phosphate and sugar groups of adjacent nucleotides. The strands are bonded to each other at their bases with hydrogen bonds, and the strands coil about each other along their length, hence the “double helix” description, which means a double spiral.

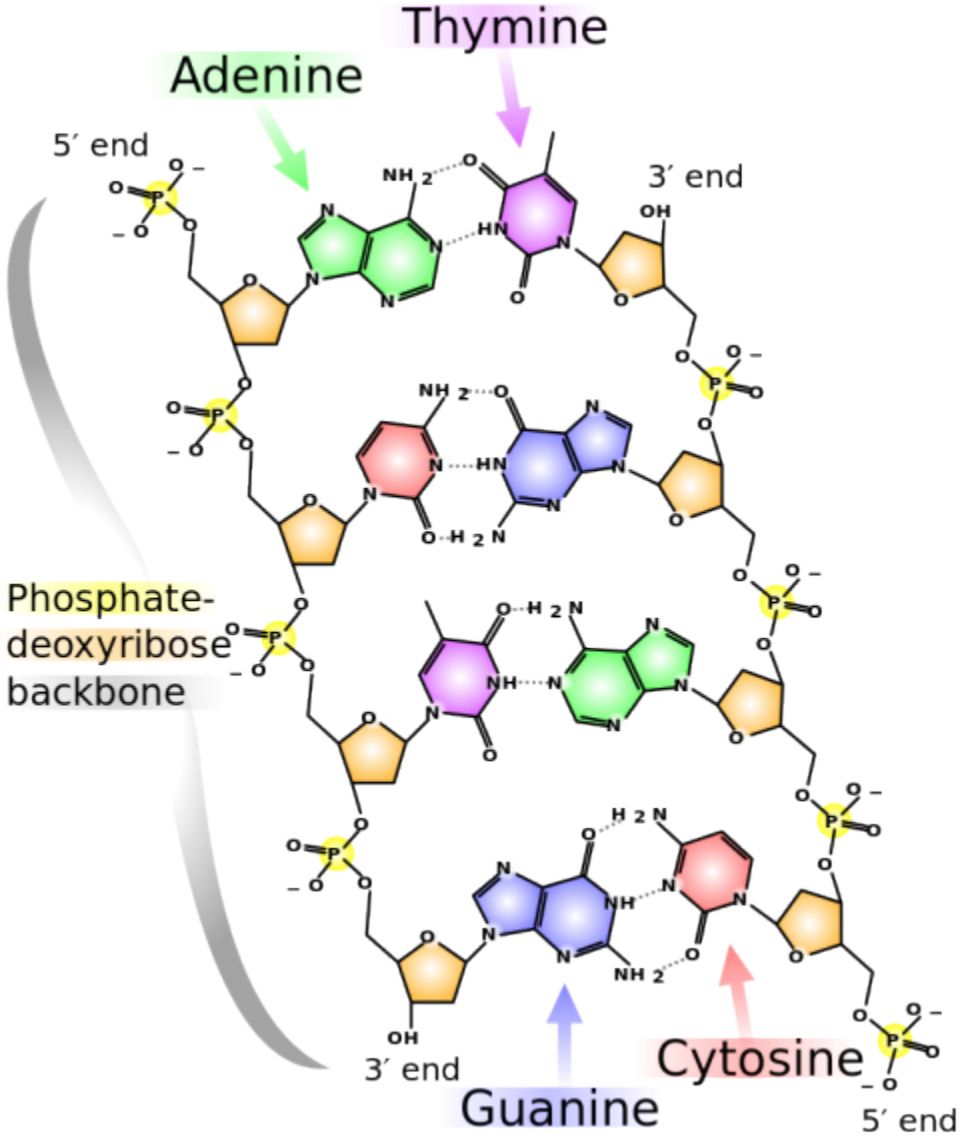


Figure 2.24 Chemical structure of DNA, with colored label identifying the four bases as well as the phosphate and deoxyribose components of the backbone.

The alternating sugar and phosphate groups lie on the outside of each strand, forming the backbone of the DNA. The nitrogenous bases are stacked in the interior, like the steps of a staircase, and these bases pair; the pairs are bound to each other by hydrogen bonds. The bases pair in such a way that the distance between the backbones of the two strands is the same all along the molecule. The rule is that nucleotide A pairs with nucleotide T, and G with C, see section 9.1 for more details.

Section Summary

Living things are carbon-based because carbon plays such a prominent role in the chemistry of living things. The four covalent bonding positions of the carbon atom can give rise to a wide diversity of compounds with many functions, accounting for the importance of carbon in living things. Carbohydrates are a group of macromolecules that are a vital energy source for the cell, provide

structural support to many organisms, and can be found on the surface of the cell as receptors or for cell recognition. Carbohydrates are classified as monosaccharides, disaccharides, and polysaccharides, depending on the number of monomers in the molecule.

Lipids are a class of macromolecules that are nonpolar and hydrophobic in nature. Major types include fats and oils, waxes, phospholipids, and steroids. Fats and oils are a stored form of energy and can include triglycerides. Fats and oils are usually made up of fatty acids and glycerol.

Proteins are a class of macromolecules that can perform a diverse range of functions for the cell. They help in metabolism by providing structural support and by acting as enzymes, carriers or as hormones. The building blocks of proteins are amino acids. Proteins are organized at four levels: primary, secondary, tertiary, and quaternary. Protein shape and function are intricately linked; any change in shape caused by changes in temperature, pH, or chemical exposure may lead to protein denaturation and a loss of function.

Nucleic acids are molecules made up of repeating units of nucleotides that direct cellular activities such as cell division and protein synthesis. Each nucleotide is made up of a pentose sugar, a nitrogenous base, and a phosphate group. There are two types of nucleic acids: DNA and RNA.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4368#h5p-15

Glossary

amino acid: a monomer of a protein

carbohydrate: a biological macromolecule in which the ratio of carbon to hydrogen to oxygen is 1:2:1; carbohydrates serve as energy sources and structural support in cells

cellulose: a polysaccharide that makes up the cell walls of plants and provides structural support to the cell

chitin: a type of carbohydrate that forms the outer skeleton of arthropods, such as insects and crustaceans, and the cell walls of fungi

denaturation: the loss of shape in a protein as a result of changes in temperature, pH, or exposure to chemicals

deoxyribonucleic acid (DNA): a double-stranded polymer of nucleotides that carries the hereditary information of the cell

disaccharide: two sugar monomers that are linked together by a peptide bond

enzyme: a catalyst in a biochemical reaction that is usually a complex or conjugated protein

fat: a lipid molecule composed of three fatty acids and a glycerol (triglyceride) that typically exists in a solid form at room temperature

glycogen: a storage carbohydrate in animals

hormone: a chemical signaling molecule, usually a protein or steroid, secreted by an endocrine gland or group of endocrine cells; acts to control or regulate specific physiological processes

lipids: a class of macromolecules that are nonpolar and insoluble in water

macromolecule: a large molecule, often formed by polymerization of smaller monomers

monosaccharide: a single unit or monomer of carbohydrates

nucleic acid: a biological macromolecule that carries the genetic information of a cell and carries instructions for the functioning of the cell

nucleotide: a monomer of nucleic acids; contains a pentose sugar, a phosphate group, and a nitrogenous base

oil: an unsaturated fat that is a liquid at room temperature

phospholipid: a major constituent of the membranes of cells; composed of two fatty acids and a phosphate group attached to the glycerol backbone

polypeptide: a long chain of amino acids linked by peptide bonds

polysaccharide: a long chain of monosaccharides; may be branched or unbranched

protein: a biological macromolecule composed of one or more chains of amino acids

ribonucleic acid (RNA): a single-stranded polymer of nucleotides that is involved in protein synthesis

saturated fatty acid: a long-chain hydrocarbon with single covalent bonds in the carbon chain; the number of hydrogen atoms attached to the carbon skeleton is maximized

starch: a storage carbohydrate in plants

steroid: a type of lipid composed of four fused hydrocarbon rings

trans-fat: a form of unsaturated fat with the hydrogen atoms neighboring the double bond across from each other rather than on the same side of the double bond

triglyceride: a fat molecule; consists of three fatty acids linked to a glycerol molecule

unsaturated fatty acid: a long-chain hydrocarbon that has one or more than one double bonds in the hydrocarbon chain

Media Attribution

by the National Human Genome Research Institute, which is in the public domain.

- Figure 2.24 by Madeleine Price Ball is licensed under a CC BY-SA 2.5 licence.

Chapter 3: Introduction to Cell Structure and Function

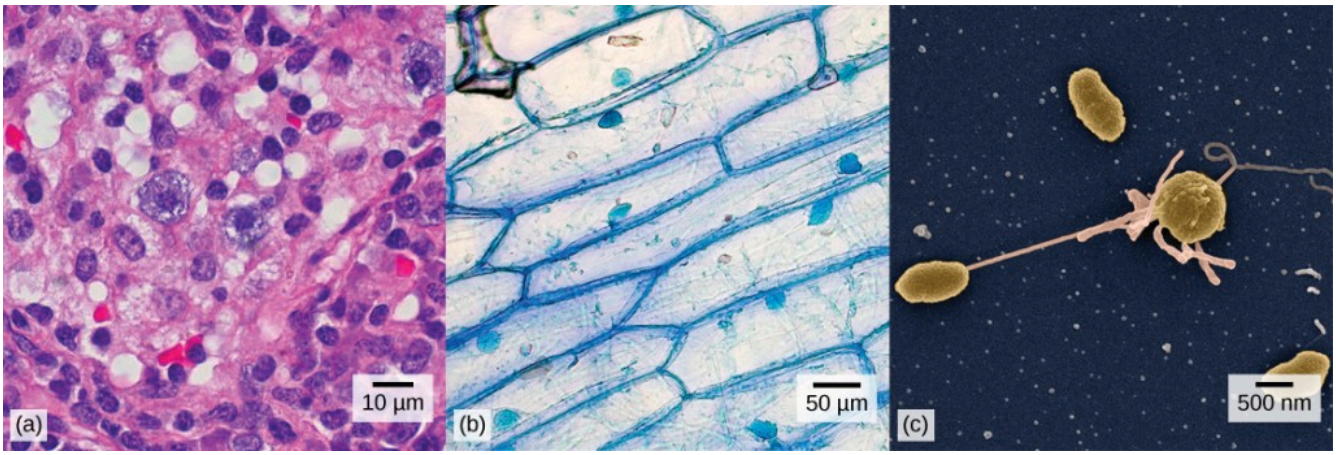


Figure 3.1 (a) Nasal sinus cells (viewed with a light microscope), (b) onion cells (viewed with a light microscope), and (c) Vibrio tasmaniensis bacterial cells (viewed using a scanning electron microscope) are from very different organisms, yet all share certain characteristics of basic cell structure. Close your eyes and picture a brick wall. What is the basic building block of that wall? It is a single brick, of course. Like a brick wall, your body is composed of basic building blocks, and the building blocks of your body are cells. An average human is thought to have 37.2 trillion cells.

Your body has many kinds of cells, each specialized for a specific purpose. Just as a home is made from a variety of building materials, the human body is constructed from many cell types. For example, epithelial cells protect the surface of the body and cover the organs and body cavities within. Bone cells help to support and protect the body. Cells of the immune system fight invading bacteria. Additionally, red blood cells carry oxygen throughout the body. Each of these cell types plays a vital role during the growth, development, and day-to-day maintenance of the body. In spite of their enormous variety, however, all cells share certain fundamental characteristics.

Search for Key Points in Chapter 3

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4371#h5p-16

Media Attribution

- Figure 3.1
    - Nasal sinus cell: modification of work by Ed Uthman, MD;
    - Onion cell: modification of work by Umberto Salvagnin;
    - Vibrio tasmaniensis bacterial cells: modification of work by Anthony D’Onofrio; scale-bar data from Matt Russell

3.1 How Cells Are Studied

Learning Objectives

By the end of this section, you will be able to:

- Describe the roles of cells in organisms
- Compare and contrast light microscopy and electron microscopy
- Summarize the cell theory

Watch a video about eukaryotic cells

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4378#h5p-17

Watch a video about diffusion

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4378#h5p-18

A cell is the smallest unit of a living thing. A living thing, like you, is called an organism. Thus, cells are the basic building blocks of all organisms.

In multicellular organisms, several cells of one particular kind interconnect with each other and perform shared functions to form tissues (for example, muscle tissue, connective tissue, and nervous tissue), several tissues combine to form an organ (for example, stomach, heart, or brain), and several organs make up an organ system (such as the digestive system, circulatory system, or nervous system). Several systems functioning together form an organism (such as an elephant, for example).

There are many types of cells, and all are grouped into one of two broad categories: prokaryotic and eukaryotic. Animal cells, plant cells, fungal cells, and protist cells are classified as eukaryotic, whereas

bacteria and archaea cells are classified as prokaryotic. Before discussing the criteria for determining whether a cell is prokaryotic or eukaryotic, let us first examine how biologists study cells.

Microscopy

Cells vary in size. With few exceptions, individual cells are too small to be seen with the naked eye, so scientists use microscopes to study them. A **microscope** is an instrument that magnifies an object. Most images of cells are taken with a microscope and are called micrographs.

Light Microscopes

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

The optics of the lenses of a light microscope changes the orientation of the image. 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. Due to the manner in which light travels through the lenses, this system of lenses produces an inverted image (binoculars and a dissecting microscope work in a similar manner, but include an additional magnification system that makes the final image appear to be upright).

Most student microscopes are classified as light microscopes (Figure 3.2 **a**). Visible light both passes through and is bent by 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 degree of enlargement of an object. Resolving power is the ability of a microscope to allow the eye to distinguish two adjacent structures as separate; the higher the resolution, the closer those two objects can be, and the better the clarity and detail of the image. When oil immersion lenses are used, magnification is usually increased to 1,000 times for the study of smaller cells, like most prokaryotic cells. Because light entering a specimen from below is focused onto the eye of an observer, the specimen can be viewed using light microscopy. For this reason, for light to pass through a specimen, the sample must be thin or translucent.

Concept in Action

For another perspective on cell size, try the HowBig interactive.

A second type of microscope used in laboratories is the dissecting microscope (Figure 3.2 **b**). These microscopes have a lower magnification (20 to 80 times the object size) than light microscopes and can provide a three-dimensional view of the specimen. Thick objects can be examined with many components in focus at the same time. These microscopes are designed to give a magnified and clear view of tissue structure as well as the anatomy of the whole organism. Like light microscopes, most modern dissecting microscopes are also binocular, meaning that they have two separate lens systems, one for each eye. The lens systems are separated by a certain distance, and therefore provide a sense of depth in the view of their subject to make manipulations by hand easier. Dissecting microscopes also have optics that correct the image so that it appears as if being seen by the naked eye and not as an inverted image. The light illuminating a sample under a dissecting microscope typically comes from above the sample, but may also be directed from below.

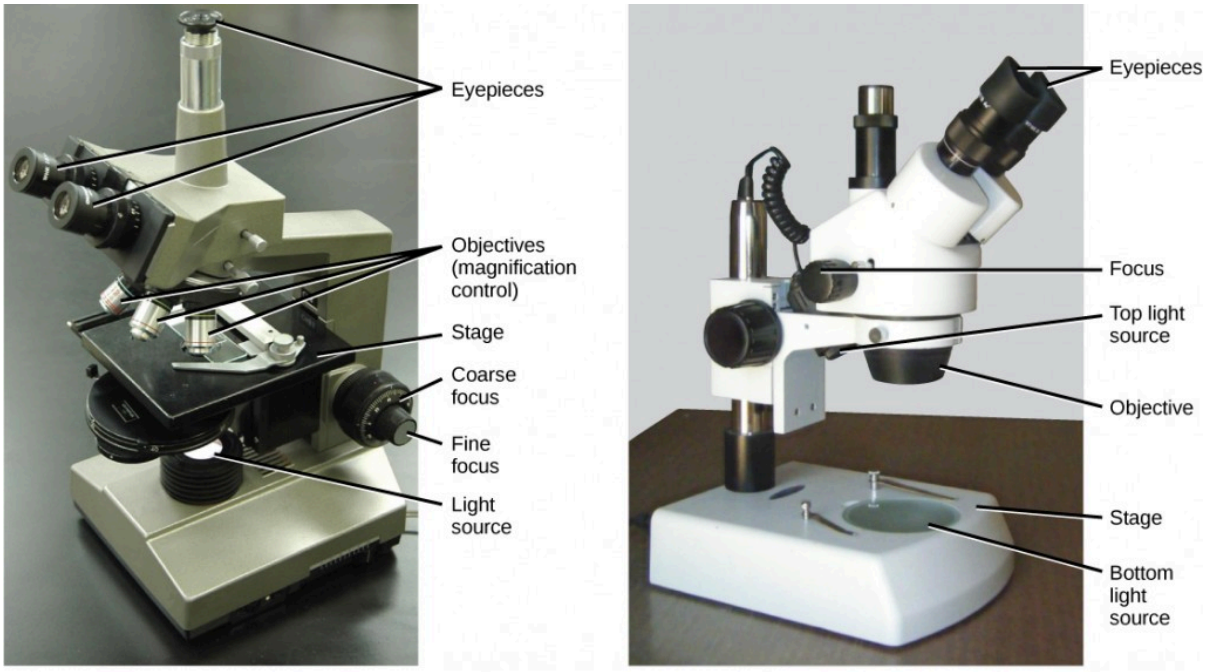


Figure 3.2 (a) Most light microscopes used in a college biology lab can magnify cells up to approximately 400 times. (b) Dissecting microscopes have a lower magnification than light microscopes and are used to examine larger objects, such as tissues.

Electron Microscopes

In contrast to light microscopes, electron microscopes use a beam of electrons instead of a beam of light. Not only does this allow for higher magnification and, thus, more detail (Figure 3.4), it also provides higher resolving power. Preparation of a specimen for viewing under an electron microscope will kill it; therefore, live cells cannot be viewed using this type of microscopy. In addition, the electron beam moves best in a vacuum, making it impossible to view living materials.

In a scanning electron microscope, a beam of electrons moves back and forth across a cell's surface, rendering the details of cell surface characteristics by reflection. Cells and other structures are usually coated with a metal like gold. In a transmission electron microscope, the electron beam is transmitted through the cell and provides details of a cell's internal structures. As you might imagine, electron microscopes are significantly more bulky and expensive than are light microscopes.

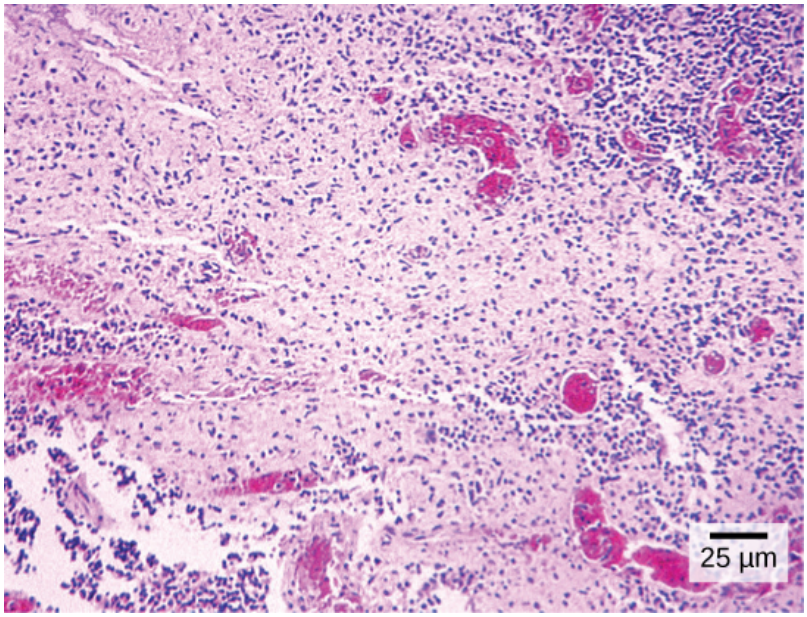


Figure 3.3 Salmonella bacteria are viewed with a light microscope.

Figure 3.4 This scanning electron micrograph shows Salmonella bacteria (in red) invading human cells.

Cytotechnologist: Have you ever heard of a medical test called a Pap smear? 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 through microscopic examinations and other laboratory tests. They are trained to determine which cellular changes are within normal limits or 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 vital roles in saving people's lives. When abnormalities are discovered early, a patient's treatment can begin sooner, which usually increases the chances of successful treatment.

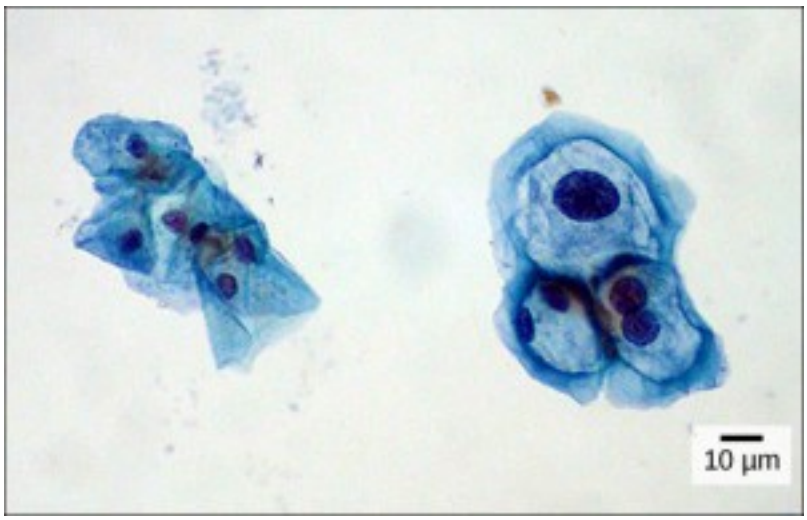


Figure 3.5 These uterine cervix cells, viewed through a light microscope, were obtained from a Pap smear. Normal cells are on the left. The cells on the right are infected with human papillomavirus.

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 protists (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” (from the Latin *cella*, meaning “small room”) 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 and microscope construction enabled other scientists to see different 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, that the cell is the basic unit of life, and that all new cells arise from existing cells. These principles still stand today.

Section Summary

A cell is the smallest unit of life. Most cells are so small that they cannot be viewed with the naked eye. Therefore, scientists must use microscopes to study cells. Electron microscopes provide higher magnification, higher resolution, and more detail than light microscopes. The unified cell theory states that all organisms are composed of one or more cells, the cell is the basic unit of life, and new cells arise from existing cells.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4378#h5p-19

Glossary

microscope: the instrument that magnifies an object

unified cell theory: the biological concept that states that all organisms are composed of one or more cells, the cell is the basic unit of life, and new cells arise from existing cells

Media Attributions

- Figure 3.3 by modification of work by CDC, Armed Forces Institute of Pathology, Charles N. Farmer © Public Domain
- Figure 3.4 modification of work by Rocky Mountain Laboratories, NIAID, NIH; scale-bar data from Matt Russell
- Figure 3.5 modification of work by Ed Uthman; scale-bar data from Matt Russell © CC BY-SA (Attribution ShareAlike)

3.2 Comparing Prokaryotic and Eukaryotic Cells

Learning Objectives

By the end of this section, you will be able to:

- Name examples of prokaryotic and eukaryotic organisms
- Compare and contrast prokaryotic cells and eukaryotic cells
- Describe the relative sizes of different kinds of cells

Cells fall into one of two broad categories: prokaryotic and eukaryotic. The predominantly single-celled organisms of the domains Bacteria and Archaea are classified as prokaryotes (*pro–* = before; *–karyon–* = nucleus). Animal cells, plant cells, fungi, and protists are eukaryotes (*eu–* = true).

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 region within the cell in which other cellular components are found; 3) DNA, the genetic material of the cell; and 4) ribosomes, particles that synthesize proteins. However, prokaryotes differ from eukaryotic cells in several ways.

A prokaryotic cell is a simple, 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 the central part of the cell: a darkened region called the nucleoid.

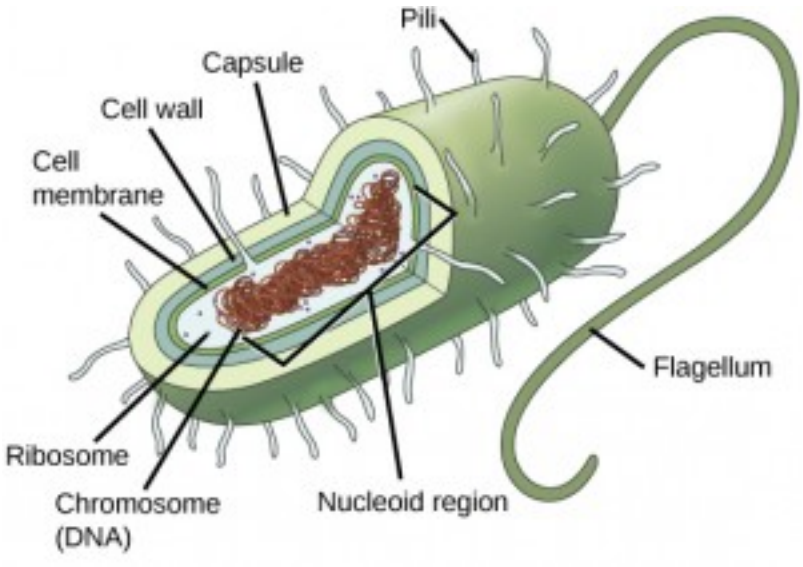


Figure 3.6 This figure shows the generalized structure of a prokaryotic cell.

Unlike Archaea and eukaryotes, bacteria have a cell wall made of peptidoglycan, comprised of sugars and amino acids, and many have a polysaccharide capsule (Figure 3.6). 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, while most pili are used to exchange genetic material during a type of reproduction called conjugation.

Eukaryotic Cells

In nature, the relationship between form and function is apparent at all levels, including the level of the cell, and this will become clear as we explore eukaryotic cells. The principle “form follows function” is found in many contexts. For example, birds and fish have streamlined bodies that allow them to move quickly through the medium in which they live, be it air or water. It means that, in general, one can deduce the function of a structure by looking at its form, because the two are matched.

A eukaryotic cell is a cell that **has a membrane-bound nucleus and other membrane-bound compartments or sacs, called organelles**, which have specialized functions. The word eukaryotic means “true kernel” or “true nucleus,” alluding to the presence of the membrane-bound nucleus in these cells. The word “organelle” means “little organ,” and, as already mentioned, organelles have specialized cellular functions, just as the organs of your body have specialized functions.

Cell Size

At 0.1–5.0 µm in diameter, prokaryotic cells are significantly smaller than eukaryotic cells, which have diameters ranging from 10–100 µm (Figure 3.7). The small size of prokaryotes allows ions and organic molecules that enter them to quickly spread to other parts of the cell. Similarly, any wastes produced within a prokaryotic cell can quickly move out. However, larger eukaryotic cells have evolved different

structural adaptations to enhance cellular transport. Indeed, the large size of these cells would not be possible without these adaptations. In general, **cell size is limited** because volume increases much more quickly than does cell surface area. As a cell becomes larger, it becomes more and more difficult for the cell to acquire sufficient materials to support the processes inside the cell, because the relative size of the surface area across which materials must be transported declines.

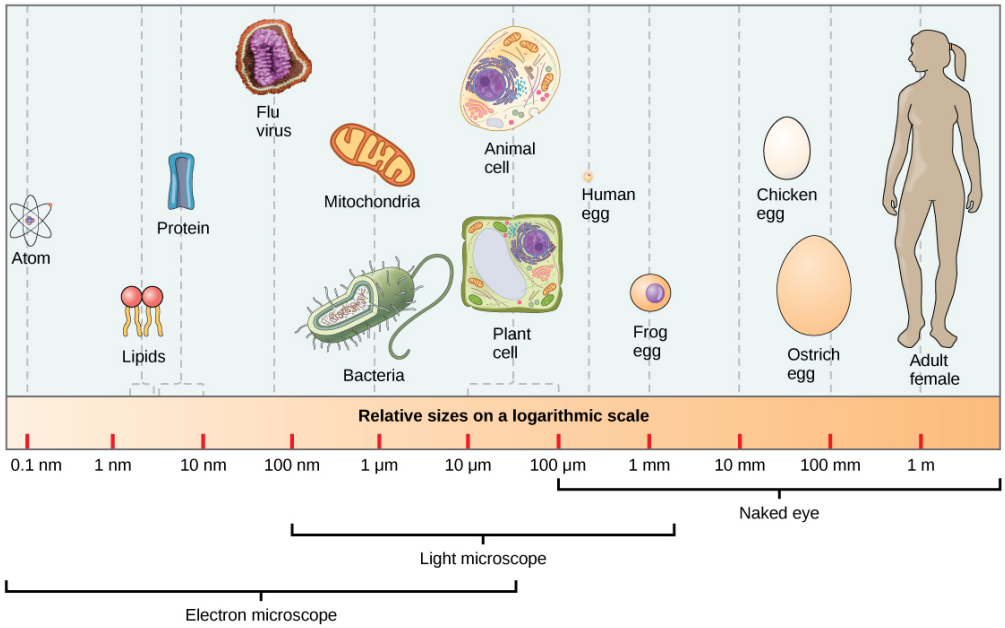


Figure 3.7 This figure shows the relative sizes of different kinds of cells and cellular components. An adult human is shown for comparison.

Section Summary

Prokaryotes are predominantly single-celled organisms of the domains Bacteria and Archaea. All prokaryotes have plasma membranes, cytoplasm, ribosomes, a cell wall, DNA, and lack membrane-bound organelles. Many also have polysaccharide capsules. Prokaryotic cells range in diameter from 0.1–5.0 µm.

Like a prokaryotic cell, a eukaryotic cell has a plasma membrane, cytoplasm, and ribosomes, but a eukaryotic cell is typically larger than a prokaryotic cell, has a true nucleus (meaning its DNA is surrounded by a membrane), and has other membrane-bound organelles that allow for compartmentalization of functions. Eukaryotic cells tend to be 10 to 100 times the size of prokaryotic cells.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4382#h5p-20

Glossary

eukaryotic cell: a cell that has a membrane-bound nucleus and several other membrane-bound compartments or sacs

organelle: a membrane-bound compartment or sac within a cell

prokaryotic cell: a unicellular organism that lacks a nucleus or any other membrane-bound organelle

3.3 Eukaryotic Cells

Learning Objectives

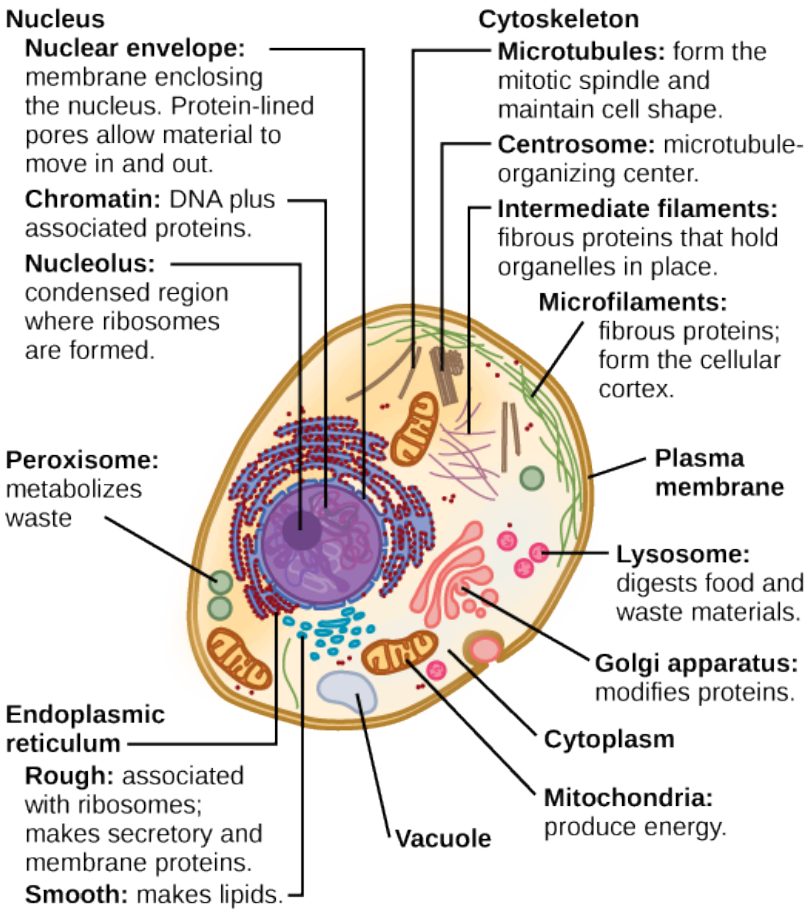
By the end of this section, you will be able to:

- Describe the structure of eukaryotic plant and animal cells
- State the role of the plasma membrane
- Summarize the functions of the major cell organelles
- Describe the cytoskeleton and extracellular matrix

Watch a video about oxygen in the atmosphere.

An interactive H5P element has been excluded from this version of the text. You can view it online here: *https://opentextbc.ca/biology/?p=4398#h5p-21*

At this point, it should be clear that eukaryotic cells have a more complex structure than do prokaryotic cells. Organelles allow for various functions to occur in the cell at the same time. Before discussing the functions of organelles within a eukaryotic cell, let us first examine two important components of the cell: the plasma membrane and the cytoplasm.



(a)

Figure 3.8 (a) This figure shows a typical animal cell

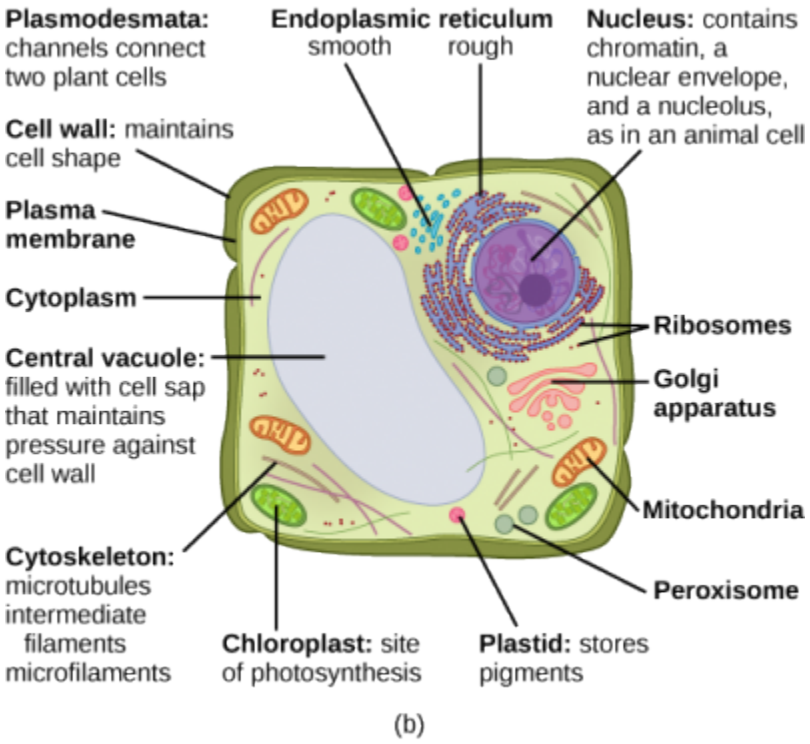


Figure 3.8 (b) This figures shows a typical plant cell.

What structures does a plant cell have that an animal cell does not have? What structures does an animal cell have that a plant cell does not have? Plant cells have plasmodesmata, a cell wall, a large central vacuole, chloroplasts, and plastids. Animal cells have lysosomes and centrosomes.

The Plasma Membrane

Like prokaryotes, eukaryotic cells have a plasma membrane (Figure 3.9) made up of a **phospholipid bilayer with embedded proteins** that separates the internal contents of the cell from its surrounding environment. A phospholipid is a lipid molecule composed of two fatty acid chains, a glycerol backbone, and a phosphate group. The plasma membrane regulates the passage of some substances, such as organic molecules, ions, and water, preventing the passage of some to maintain internal conditions, while actively bringing in or removing others. Other compounds move passively across the membrane.

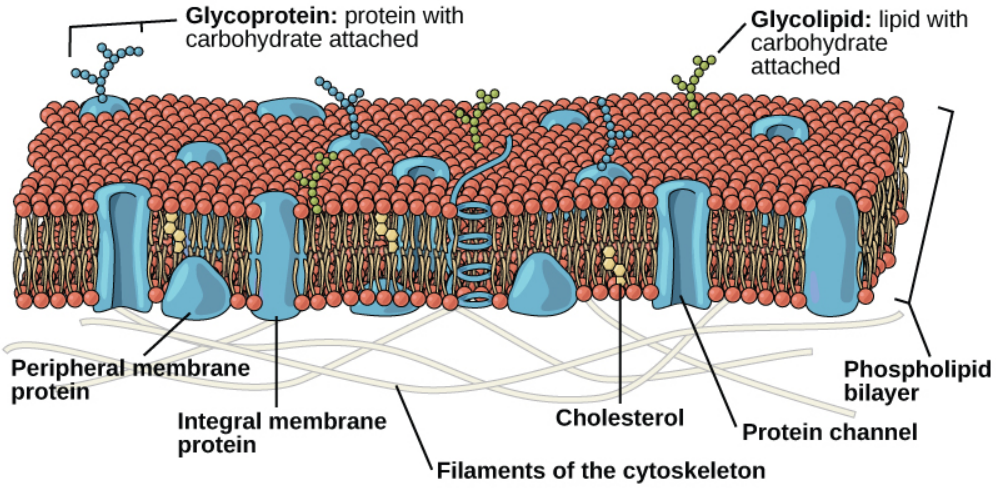


Figure 3.9 The plasma membrane is a phospholipid bilayer with embedded proteins. There are other components, such as cholesterol and carbohydrates, which can be found in the membrane in addition to phospholipids and protein.

The plasma membranes of cells that specialize in absorption are folded into fingerlike projections called microvilli (singular = microvillus). This folding increases the surface area of the plasma membrane. Such cells are typically found lining the small intestine, the organ that absorbs nutrients from digested food. This is an excellent example of form matching the function of a structure.

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.

The Cytoplasm

The cytoplasm comprises the contents 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. 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 Cytoskeleton

If you were to remove all the organelles from a cell, would the plasma membrane and the cytoplasm be the only components left? No. Within the cytoplasm, there would still be ions and organic molecules, plus a **network of protein fibers** that helps to maintain the shape of the cell, secures certain organelles in specific positions, allows cytoplasm and vesicles to move within the cell, and enables unicellular organisms to move independently. Collectively, this network of protein fibers is known as the

cytoskeleton. There are three types of fibers within the cytoskeleton: microfilaments, also known as actin filaments, intermediate filaments, and microtubules (Figure 3.10).

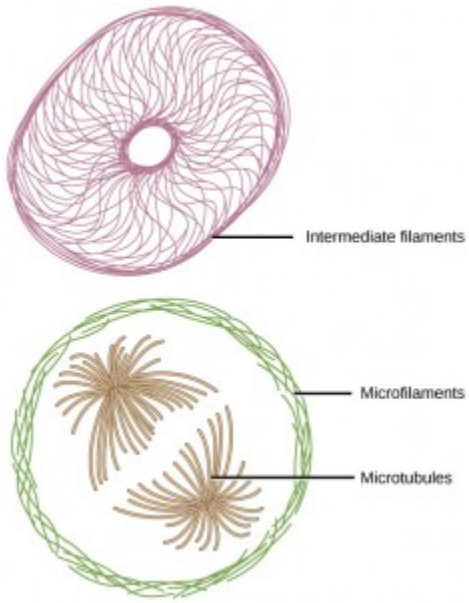


Figure 3.10 Microfilaments, intermediate filaments, and microtubules compose a cell's cytoskeleton.

Microfilaments are the thinnest of the cytoskeletal fibers and function in moving cellular components, for example, during cell division. They also maintain the structure of microvilli, the extensive folding of the plasma membrane found in cells dedicated to absorption. These components are also common in muscle cells and are responsible for muscle cell contraction. Intermediate filaments are of intermediate diameter and have structural functions, such as maintaining the shape of the cell and anchoring organelles. Keratin, the compound that strengthens hair and nails, forms one type of intermediate filament. Microtubules are the thickest of the cytoskeletal fibers. These are hollow tubes that can dissolve and reform quickly. Microtubules guide organelle movement and are the structures that pull chromosomes to their poles during cell division. They are also the structural components of flagella and cilia. In cilia and flagella, the microtubules are organized as a circle of nine double microtubules on the outside and two microtubules in the center.

The centrosome is a region near the nucleus of animal cells that functions as a microtubule-organizing center. It contains a pair of centrioles, two structures that lie perpendicular to each other. Each centriole is a cylinder of nine triplets of microtubules.

The centrosome replicates itself before a cell divides, and the centrioles play a role in pulling the duplicated chromosomes to opposite ends of the dividing cell. However, the exact function of the centrioles in cell division is not clear, since cells that have the centrioles removed can still divide, and plant cells, which lack centrioles, are capable of cell division.

Flagella and Cilia

Flagella (singular = flagellum) are long, hair-like structures that extend from the plasma membrane and are used to move an entire cell, (for example, sperm, *Euglena*). When present, the cell has just one flagellum or a few flagella. When cilia (singular = cilium) are present, however, they are many in number and extend along the entire surface of the plasma membrane. They are short, hair-like structures that are used to move entire cells (such as paramecium) or move substances along the outer surface of the cell (for example, the cilia of cells lining the fallopian tubes that move the ovum toward the uterus, or cilia lining the cells of the respiratory tract that move particulate matter toward the throat that mucus has trapped).

The Endomembrane System

The endomembrane system (*endo* = within) is a **group of membranes and organelles in eukaryotic cells that work together to modify, package, and transport lipids and proteins**. It includes the nuclear envelope, lysosomes, vesicles, endoplasmic reticulum and the Golgi apparatus, which we will cover shortly. Although not technically *within* the cell, the plasma membrane is included in the endomembrane system because, as you will see, it interacts with the other endomembranous organelles.

The Nucleus

Typically, the nucleus is the most prominent organelle in a cell. The nucleus (plural = nuclei) **houses the cell's DNA** in the form of chromatin and directs the synthesis of ribosomes and proteins. Let us look at it in more detail (Figure 3.11).

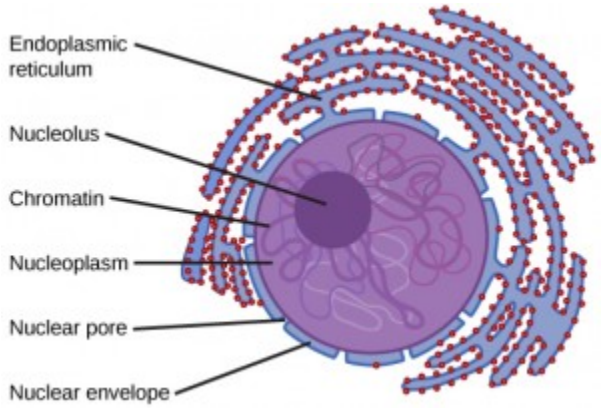


Figure 3.11 The outermost boundary of the nucleus is the nuclear envelope. Notice that the nuclear envelope consists of two phospholipid bilayers (membranes)—an outer membrane and an inner membrane—in contrast to the plasma membrane, which consists of only one phospholipid bilayer.

The nuclear envelope is a **double-membrane structure** that constitutes the outermost portion of the

nucleus (Figure 3.11). 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 the cytoplasm.

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, and proteins. This combination of DNA and proteins is called chromatin. In eukaryotes, chromosomes are linear structures. Every species has a specific number of chromosomes in the nucleus of its body cells. For example, in humans, the chromosome number is 46, whereas in fruit flies, the chromosome number 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, the chromosomes resemble an unwound, jumbled bunch of threads.

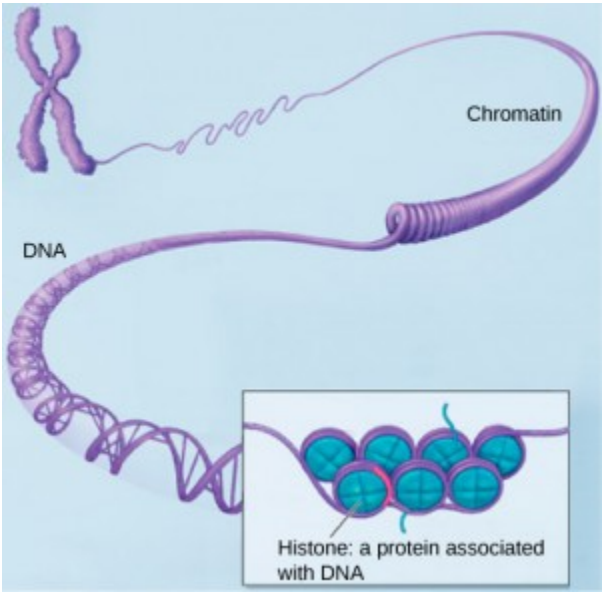


Figure 3.12 This image shows various levels of the organization of chromatin (DNA and protein).

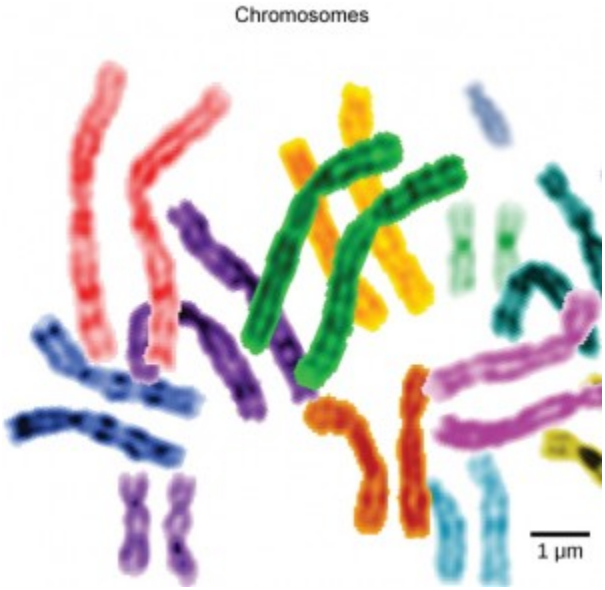


Figure 3.13 This image shows paired chromosomes. (credit: modification of work by NIH; scale-bar data from Matt Russell)

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 stained 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 through the nuclear pores into the cytoplasm.

The Endoplasmic Reticulum

The endoplasmic reticulum (ER) is a series of interconnected membranous tubules that collectively modify proteins and synthesize lipids. However, these two functions are performed in separate areas of the endoplasmic reticulum: the rough endoplasmic reticulum and the smooth endoplasmic reticulum, respectively.

The hollow portion of the ER tubules is called the lumen or cisternal space. The membrane of the ER, which is a phospholipid bilayer embedded with proteins, is continuous with the nuclear envelope.

The **rough endoplasmic reticulum (RER)** is so named because the ribosomes attached to its cytoplasmic surface give it a studded appearance when viewed through an electron microscope.

The ribosomes synthesize proteins while attached to the ER, resulting in the transfer of their newly synthesized proteins into the lumen of the RER where they undergo modifications such as folding or addition of sugars. The RER also makes phospholipids for cell membranes.

If the phospholipids or modified proteins are not destined to stay in the RER, they will be packaged within vesicles and transported from the RER by budding from the membrane. Since the RER is engaged in modifying proteins that will be secreted from the cell, it is abundant in cells that secrete proteins, such as the liver.

The **smooth endoplasmic reticulum (SER)** is continuous with the RER but has few or no ribosomes on its cytoplasmic surface. The SER's functions include synthesis of carbohydrates, lipids (including phospholipids), and steroid hormones; detoxification of medications and poisons; alcohol metabolism; and storage of calcium ions.

The Golgi Apparatus

We have already mentioned that vesicles can bud from the ER, but where do the vesicles go? Before reaching their final destination, the lipids or proteins within the transport vesicles need to be sorted, packaged, and tagged so that they wind up in the right place. The **sorting, tagging, packaging, and distribution of lipids and proteins** take place in the Golgi apparatus (also called the Golgi body), a series of flattened membranous sacs.

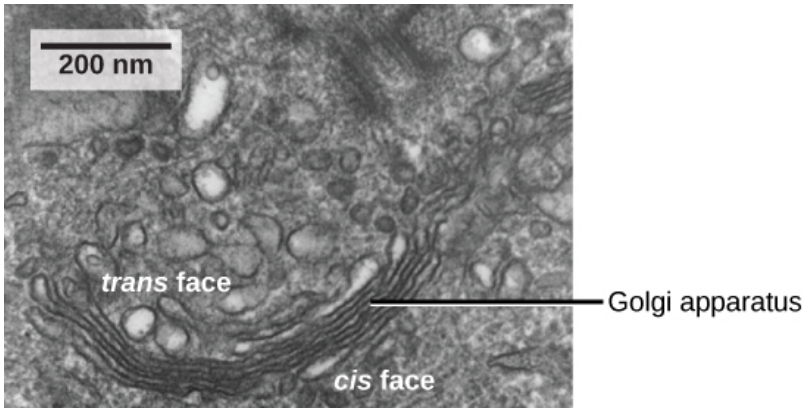


Figure 3.14 The Golgi apparatus in this transmission electron micrograph of a white blood cell is visible as a stack of semicircular flattened rings in the lower portion of this image. Several vesicles can be seen near the Golgi apparatus. (credit: modification of work by Louisa Howard; scale-bar data from Matt Russell)

The Golgi apparatus has a receiving face near the endoplasmic reticulum and a releasing face on the side away from the ER, toward the cell membrane. The transport vesicles that form from the ER travel to the receiving face, fuse with it, and empty their contents into the lumen of the Golgi apparatus. As the proteins and lipids travel through the Golgi, they undergo further modifications. The most frequent modification is the addition of short chains of sugar molecules. The newly modified proteins and lipids are then tagged with small molecular groups to enable them to be routed to their proper destinations.

Finally, the modified and tagged proteins are packaged into vesicles that bud from the opposite face of the Golgi. While some of these vesicles, transport vesicles, deposit their contents into other parts of the cell where they will be used, others, secretory vesicles, fuse with the plasma membrane and release their contents outside the cell.

The amount of Golgi in different cell types again illustrates that form follows function within cells. Cells that engage in a great deal of secretory activity (such as cells of the salivary glands that secrete digestive enzymes or cells of the immune system that secrete antibodies) have an abundant number of Golgi.

In plant cells, the Golgi has an additional role of synthesizing polysaccharides, some of which are incorporated into the cell wall and some of which are used in other parts of the cell.

Lysosomes

In animal cells, the lysosomes are the cell's **"garbage disposal."** Digestive enzymes within the lysosomes aid the breakdown of proteins, polysaccharides, lipids, nucleic acids, and even worn-out organelles. In single-celled eukaryotes, lysosomes are important for digestion of the food they ingest and the **recycling of organelles**. These enzymes are active at a much lower pH (more acidic) than those located in the cytoplasm. Many reactions that take place in the cytoplasm could not occur at a low pH, thus the advantage of compartmentalizing the eukaryotic cell into organelles is apparent.

Lysosomes also use their hydrolytic enzymes to destroy disease-causing organisms that might enter the cell. A good example of this occurs in a group of white blood cells called macrophages, which are part of your body's immune system. In a process known as phagocytosis, a section of the plasma membrane of the macrophage invaginates (folds in) and engulfs a pathogen. The invaginated section, with the pathogen inside, then pinches itself off from the plasma membrane and becomes a vesicle. The vesicle fuses with a lysosome. The lysosome's hydrolytic enzymes then destroy the pathogen (Figure 3.15).

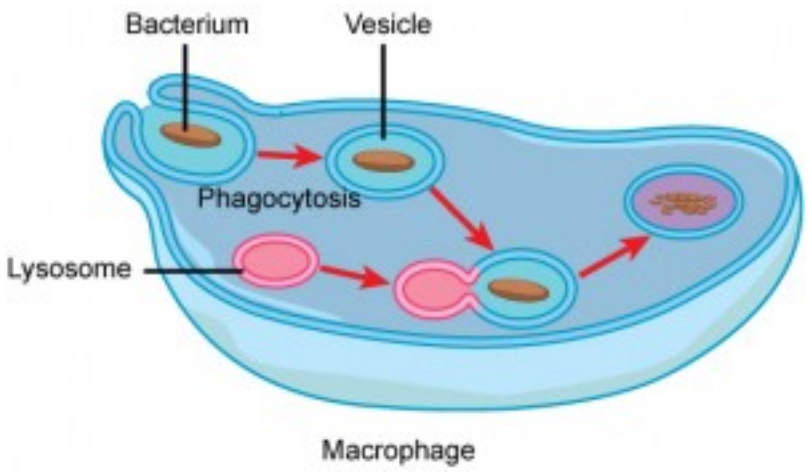


Figure 3.15 A macrophage has phagocytized a potentially pathogenic bacterium into a vesicle, which then fuses with a lysosome within the cell so that the pathogen can be destroyed. Other organelles are present in the cell, but for simplicity, are not shown.

Vesicles and Vacuoles

Vesicles and vacuoles are membrane-bound sacs that function in storage and transport. Vacuoles are somewhat larger than vesicles, and the membrane of a vacuole does not fuse with the membranes of other cellular components. Vesicles can fuse with other membranes within the cell system. Additionally, enzymes within plant vacuoles can break down macromolecules.

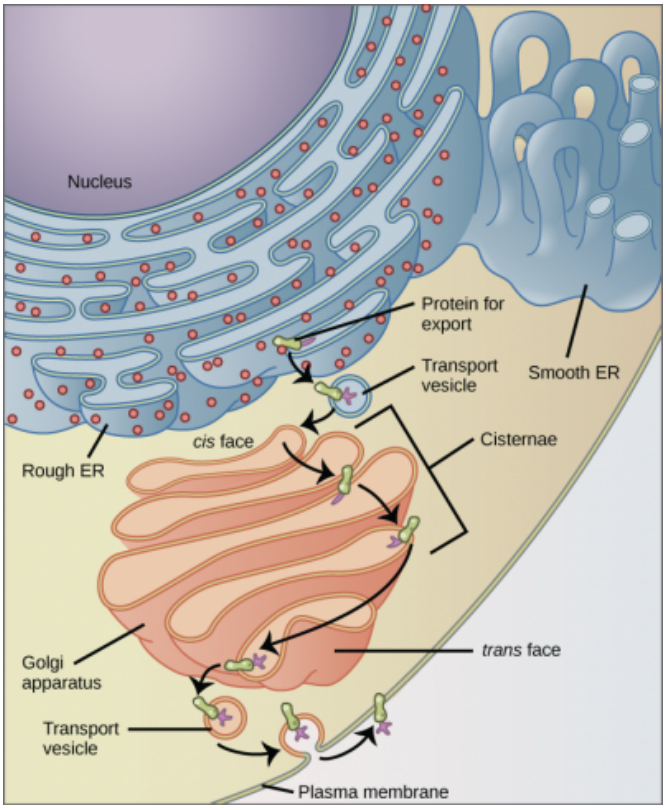


Figure 3.16 The endomembrane system works to modify, package, and transport lipids and proteins.

Why does the *cis* face of the Golgi not face the plasma membrane?

<!– Because that face receives chemicals from the ER, which is toward the center of the cell. –>

Ribosomes

Ribosomes are the cellular structures responsible for **protein synthesis**. When viewed through an electron microscope, free ribosomes appear as either clusters or single tiny dots floating freely in the cytoplasm. Ribosomes may be attached to either the cytoplasmic side of the plasma membrane or the cytoplasmic side of the endoplasmic reticulum. Electron microscopy has shown that ribosomes consist of large and small subunits. Ribosomes are enzyme complexes that are responsible for protein synthesis.

Because protein synthesis is essential for all cells, ribosomes are found in practically every cell, although they are smaller in prokaryotic cells. They are particularly abundant in immature red blood cells for the synthesis of hemoglobin, which functions in the transport of oxygen throughout the body.

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. The **formation of ATP** from the breakdown of glucose is known as cellular respiration. Mitochondria are oval-shaped, double-membrane organelles (Figure 3.17) that have their own ribosomes and DNA. Each membrane is a phospholipid bilayer embedded with proteins. The inner layer has folds called cristae, which increase the surface area of the inner membrane. The area surrounded by the folds is called the mitochondrial matrix. The cristae and the matrix have different roles in cellular respiration.

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 because muscle cells need a lot of energy to contract.

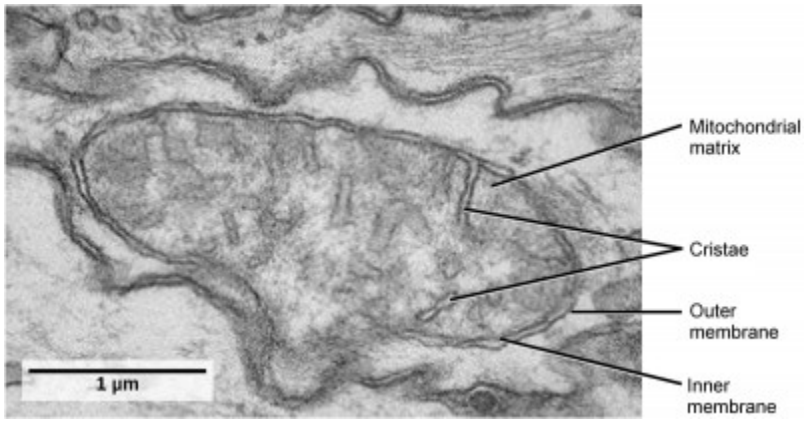


Figure 3.17 This transmission electron micrograph shows a mitochondrion as viewed with an electron microscope. Notice the inner and outer membranes, the cristae, and the mitochondrial matrix.

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. Alcohol is detoxified by peroxisomes in liver cells. A byproduct of these oxidation reactions is hydrogen peroxide, H_2O_2, which is contained within the peroxisomes to prevent the chemical from causing damage to cellular components outside of the organelle. Hydrogen peroxide is safely broken down by peroxisomal enzymes into water and oxygen.

Animal Cells versus Plant Cells

Despite their fundamental similarities, there are some striking differences between animal and plant cells (see Table 3.1). Animal cells have centrioles, centrosomes (discussed under the cytoskeleton), and lysosomes, whereas plant cells do not. Plant cells have a cell wall, chloroplasts, plasmodesmata, and plastids used for storage, and a large central vacuole, whereas animal cells do not.

The Cell Wall

In Figure 3.8 **b**, the diagram of a plant cell, you 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 protist 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, a polysaccharide made up of long, straight chains of glucose units. When nutritional information refers to dietary fiber, it is referring to the cellulose content of food.

Chloroplasts

Like mitochondria, chloroplasts also have their own DNA and ribosomes. Chloroplasts function in photosynthesis and can be found in eukaryotic cells such as plants and algae. In photosynthesis, carbon dioxide, water, and light energy are used to make glucose and oxygen. This is the major difference between plants and animals: Plants (autotrophs) are able to make their own food, like glucose, whereas animals (heterotrophs) must rely on other organisms for their organic compounds or food source.

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 3.18). Each stack of thylakoids is called a granum (plural = grana). The fluid enclosed by the inner membrane and surrounding the grana is called the stroma.

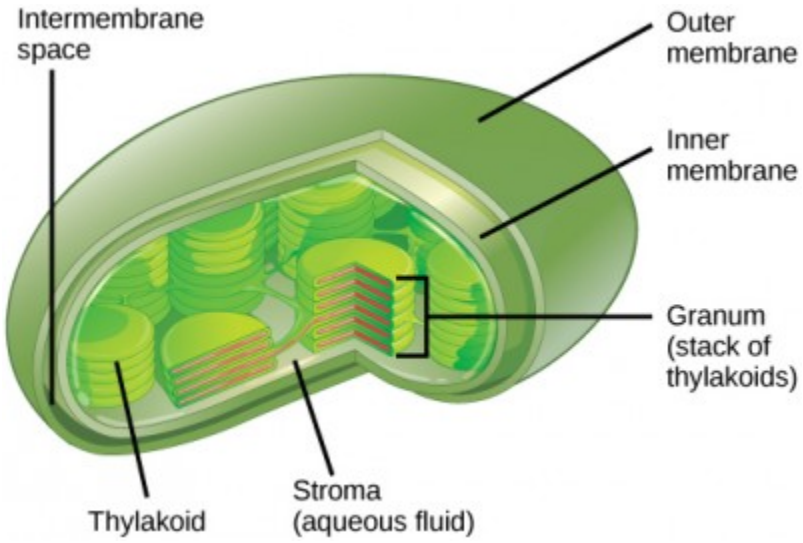


Figure 3.18 This simplified diagram of a chloroplast shows the outer membrane, inner membrane, thylakoids, grana, and stroma.

The chloroplasts contain a green pigment called chlorophyll, which captures the energy of sunlight for photosynthesis. Like plant cells, photosynthetic protists also have chloroplasts. Some bacteria also perform photosynthesis, but they do not have chloroplasts. Their photosynthetic pigments are located in the thylakoid membrane within the cell itself.

Evolution in Action

Endosymbiosis: We have mentioned that both mitochondria and chloroplasts contain DNA and ribosomes. Have you wondered why? Strong evidence points to endosymbiosis as the explanation.

Symbiosis is a relationship in which organisms from two separate species live in close association and typically exhibit specific adaptations to each other. Endosymbiosis (*endo*-= within) is a relationship in which one organism lives inside the other. Endosymbiotic relationships abound in nature. Microbes that produce vitamin K live inside the human gut. This relationship is beneficial for us because we are unable to synthesize vitamin K. It is also beneficial for the microbes because they are protected from other organisms and are provided a stable habitat and abundant food by living within the large intestine.

Scientists have long noticed that bacteria, mitochondria, and chloroplasts are similar in size. We also know that mitochondria and chloroplasts have DNA and ribosomes, just as bacteria do and they resemble the types found in bacteria. Scientists believe that host cells and bacteria formed a mutually beneficial endosymbiotic relationship when the host cells ingested aerobic bacteria and cyanobacteria but did not destroy them. Through evolution, these ingested bacteria became more specialized in their functions, with the aerobic bacteria becoming mitochondria and the photosynthetic bacteria becoming chloroplasts.

The Central Vacuole

Previously, we mentioned vacuoles as essential components of plant cells. If you look at Figure 3.8 **b**, you will see that plant cells each have a large, central vacuole that occupies most of the cell. The central vacuole plays a key role in regulating the cell's concentration of water in changing environmental conditions. In plant cells, the liquid inside the central vacuole provides turgor pressure, which is the outward pressure caused by the fluid inside the cell. Have you ever noticed that if you forget to water a plant for a few days, it wilts? That is 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 and into the soil. As the central vacuole shrinks, it leaves the cell wall unsupported. This loss of support to the cell walls of a plant results in the wilted appearance. Additionally, this fluid has a very bitter taste, which discourages consumption by insects and animals. The central vacuole also functions to store proteins in developing seed cells.

Extracellular Matrix of Animal Cells

Most animal cells release materials into the extracellular space. The primary components of these materials are glycoproteins and the protein collagen. Collectively, these materials are called the extracellular matrix (Figure 3.19). Not only does the extracellular matrix hold the cells together to form a tissue, but it also allows the cells within the tissue to communicate with each other.

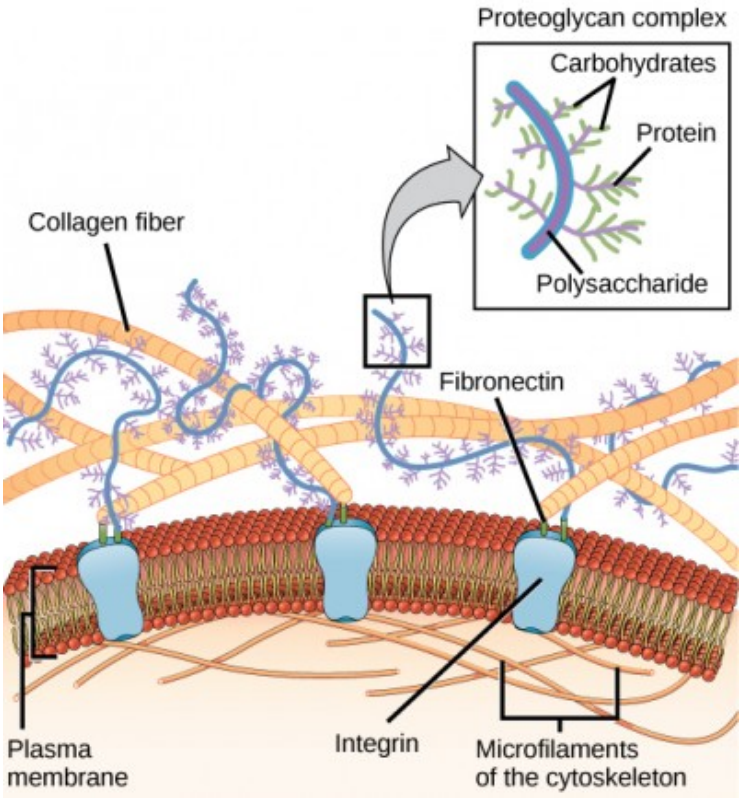


Figure 3.19 The extracellular matrix consists of a network of substances secreted by cells.

Blood clotting provides an example of the role of the extracellular matrix in cell communication. When the cells lining a blood vessel are damaged, they display a protein receptor called tissue factor. When tissue factor binds with another factor in the extracellular matrix, it causes platelets to adhere to the wall of the damaged blood vessel, stimulates adjacent smooth muscle cells in the blood vessel to contract (thus constricting the blood vessel), and initiates a series of steps that stimulate the platelets to produce clotting factors.

Intercellular Junctions

Cells can also communicate with each other by direct contact, referred to as intercellular junctions. There are some differences in the ways that plant and animal cells do this. Plasmodesmata (singular = plasmodesma) are junctions between plant cells, whereas animal cell contacts include tight and gap junctions, and desmosomes.

In general, long stretches of the plasma membranes of neighboring plant cells cannot touch one another because they are separated by the cell walls surrounding each cell. Plasmodesmata are numerous channels that pass between the cell walls of adjacent plant cells, connecting their cytoplasm and enabling signal molecules and nutrients to be transported from cell to cell (Figure 3.20 **a**).

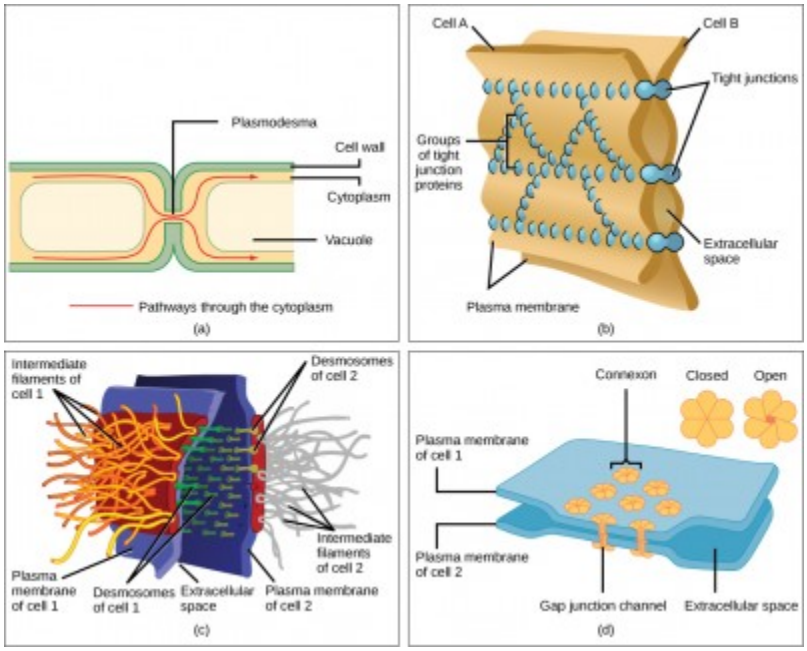


Figure 3.20 There are four kinds of connections between cells. (a) A plasmodesma is a channel between the cell walls of two adjacent plant cells. (b) Tight junctions join adjacent animal cells. (c) Desmosomes join two animal cells together. (d) Gap junctions act as channels between animal cells.

A tight junction is a watertight seal between two adjacent animal cells (Figure 3.20 **b**). Proteins hold the cells tightly against each other. This tight adhesion prevents materials from leaking between the cells. Tight junctions are typically found in the epithelial tissue that lines internal organs and cavities, and composes most of the skin. For example, the tight junctions of the epithelial cells lining the urinary bladder prevent urine from leaking into the extracellular space.

Also found only in animal cells are desmosomes, which act like spot welds between adjacent epithelial cells (Figure 3.20 **c**). They keep cells together in a sheet-like formation in organs and tissues that stretch, like the skin, heart, and muscles.

Gap junctions in animal cells are like plasmodesmata in plant cells in that they are channels between adjacent cells that allow for the transport of ions, nutrients, and other substances that enable cells to communicate (Figure 3.20 **d**). Structurally, however, gap junctions and plasmodesmata differ.

Table 3.1 Components of Prokaryotic and Eukaryotic Cells and Their Functions

Cell Component	Function	Present in Prokaryotes?	Present in Animal Cells?	Present in Plant Cells?
Plasma membrane	Separates cell from external environment; controls passage of organic molecules, ions, water, oxygen, and wastes into and out of the cell	Yes	Yes	Yes
Cytoplasm	Provides structure to cell; site of many metabolic reactions; medium in which organelles are found	Yes	Yes	Yes
Nucleoid	Location of DNA	Yes	No	No
Nucleus	Cell organelle that houses DNA and directs synthesis of ribosomes and proteins	No	Yes	Yes
Ribosomes	Protein synthesis	Yes	Yes	Yes
Mitochondria	ATP production/cellular respiration	No	Yes	Yes
Peroxisomes	Oxidizes and breaks down fatty acids and amino acids, and detoxifies poisons	No	Yes	Yes
Vesicles and vacuoles	Storage and transport; digestive function in plant cells	No	Yes	Yes
Centrosome	Unspecified role in cell division in animal cells; organizing center of microtubules in animal cells	No	Yes	No
Lysosomes	Digestion of macromolecules; recycling of worn-out organelles	No	Yes	No
Cell wall	Protection, structural support and maintenance of cell shape	Yes, primarily peptidoglycan in bacteria but not Archaea	No	Yes, primarily cellulose
Chloroplasts	Photosynthesis	No	No	Yes
Endoplasmic reticulum	Modifies proteins and synthesizes lipids	No	Yes	Yes
Golgi apparatus	Modifies, sorts, tags, packages, and distributes lipids and proteins	No	Yes	Yes
Cytoskeleton	Maintains cell's shape, secures organelles in specific positions, allows cytoplasm and vesicles to move within the cell, and enables unicellular organisms to move independently	Yes	Yes	Yes

Cell Component	Function	Present in Prokaryotes?	Present in Animal Cells?	Present in Plant Cells?
Flagella	Cellular locomotion	Some	Some	No, except for some plant sperm.
Cilia	Cellular locomotion, movement of particles along extracellular surface of plasma membrane, and filtration	No	Some	No

Section Summary

Like a prokaryotic cell, a eukaryotic cell has a plasma membrane, cytoplasm, and ribosomes, but a eukaryotic cell is typically larger than a prokaryotic cell, has a true nucleus (meaning its DNA is surrounded by a membrane), and has other membrane-bound organelles that allow for compartmentalization of functions. The plasma membrane is a phospholipid bilayer embedded with proteins. The nucleolus within the nucleus is the site for ribosome assembly. Ribosomes are found in the cytoplasm or are attached to the cytoplasmic side of the plasma membrane or endoplasmic reticulum. They perform protein synthesis. Mitochondria perform cellular respiration and produce ATP. Peroxisomes break down fatty acids, amino acids, and some toxins. Vesicles and vacuoles are storage and transport compartments. In plant cells, vacuoles also help break down macromolecules.

Animal cells also have a centrosome and lysosomes. The centrosome has two bodies, the centrioles, with an unknown role in cell division. Lysosomes are the digestive organelles of animal cells.

Plant cells have a cell wall, chloroplasts, and a central vacuole. The plant cell wall, whose primary component is cellulose, protects the cell, provides structural support, and gives shape to the cell. Photosynthesis takes place in chloroplasts. The central vacuole expands, enlarging the cell without the need to produce more cytoplasm.

The endomembrane system includes the nuclear envelope, the endoplasmic reticulum, Golgi apparatus, lysosomes, vesicles, as well as the plasma membrane. These cellular components work together to modify, package, tag, and transport membrane lipids and proteins.

The cytoskeleton has three different types of protein elements. Microfilaments provide rigidity and shape to the cell, and facilitate cellular movements. Intermediate filaments bear tension and anchor the nucleus and other organelles in place. Microtubules help the cell resist compression, serve as tracks for motor proteins that move vesicles through the cell, and pull replicated chromosomes to opposite ends of a dividing cell. They are also the structural elements of centrioles, flagella, and cilia.

Animal cells communicate through their extracellular matrices and are connected to each other by tight

junctions, desmosomes, and gap junctions. Plant cells are connected and communicate with each other by plasmodesmata.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4398#h5p-22

Glossary

cell wall: a rigid cell covering made of cellulose in plants, peptidoglycan in bacteria, non-peptidoglycan compounds in Archaea, and chitin in fungi that protects the cell, provides structural support, and gives shape to the cell

central vacuole: a large plant cell organelle that acts as a storage compartment, water reservoir, and site of macromolecule degradation

chloroplast: a plant cell organelle that carries out photosynthesis

cilium: (plural: cilia) a short, hair-like structure that extends from the plasma membrane in large numbers and is used to move an entire cell or move substances along the outer surface of the cell

cytoplasm: the entire region between the plasma membrane and the nuclear envelope, consisting of organelles suspended in the gel-like cytosol, the cytoskeleton, and various chemicals

cytoskeleton: the network of protein fibers that collectively maintains the shape of the cell, secures some organelles in specific positions, allows cytoplasm and vesicles to move within the cell, and enables unicellular organisms to move

cytosol: the gel-like material of the cytoplasm in which cell structures are suspended

desmosome: a linkage between adjacent epithelial cells that forms when cadherins in the plasma membrane attach to intermediate filaments

endomembrane system: the group of organelles and membranes in eukaryotic cells that work together to modify, package, and transport lipids and proteins

endoplasmic reticulum (ER): a series of interconnected membranous structures within eukaryotic cells that collectively modify proteins and synthesize lipids

extracellular matrix: the material, primarily collagen, glycoproteins, and proteoglycans, secreted from animal cells that holds cells together as a tissue, allows cells to communicate with each other, and provides mechanical protection and anchoring for cells in the tissue

flagellum: (plural: flagella) the long, hair-like structure that extends from the plasma membrane and is used to move the cell

gap junction: a channel between two adjacent animal cells that allows ions, nutrients, and other low-molecular weight substances to pass between the cells, enabling the cells to communicate

Golgi apparatus: a eukaryotic organelle made up of a series of stacked membranes that sorts, tags, and packages lipids and proteins for distribution

lysosome: an organelle in an animal cell that functions as the cell's digestive component; it breaks down proteins, polysaccharides, lipids, nucleic acids, and even worn-out organelles

mitochondria: (singular: mitochondrion) the cellular organelles responsible for carrying out cellular respiration, resulting in the production of ATP, the cell's main energy-carrying molecule

nuclear envelope: the double-membrane structure that constitutes the outermost portion of the nucleus

nucleolus: the darkly staining body within the nucleus that is responsible for assembling ribosomal subunits

nucleus: the cell organelle that houses the cell's DNA and directs the synthesis of ribosomes and proteins

peroxisome: a small, round organelle that contains hydrogen peroxide, oxidizes fatty acids and amino acids, and detoxifies many poisons

plasma membrane: a phospholipid bilayer with embedded (integral) or attached (peripheral) proteins that separates the internal contents of the cell from its surrounding environment

plasmodesma: (plural: plasmodesmata) a channel that passes between the cell walls of adjacent plant cells, connects their cytoplasm, and allows materials to be transported from cell to cell

ribosome: a cellular structure that carries out protein synthesis

rough endoplasmic reticulum (RER): the region of the endoplasmic reticulum that is studded with ribosomes and engages in protein modification

smooth endoplasmic reticulum (SER): the region of the endoplasmic reticulum that has few or no ribosomes on its cytoplasmic surface and synthesizes carbohydrates, lipids, and steroid hormones; detoxifies chemicals like pesticides, preservatives, medications, and environmental pollutants, and stores calcium ions

tight junction: a firm seal between two adjacent animal cells created by protein adherence

vacuole: a membrane-bound sac, somewhat larger than a vesicle, that functions in cellular storage and transport

vesicle: a small, membrane-bound sac that functions in cellular storage and transport; its membrane is capable of fusing with the plasma membrane and the membranes of the endoplasmic reticulum and Golgi apparatus

Media Attribution

- Figure 3.11: modification of work by NIGMS, NIH
- Figure 3.13: modification of work by NIH; scale-bar data from Matt Russell
- Figure 3.14: modification of work by Louisa Howard; scale-bar data from Matt Russell
- Figure 3.16: modification of work by Magnus Manske
- Figure 3.17: modification of work by Matthew Britton; scale-bar data from Matt Russell
- Figure 3.20: modification of work by Mariana Ruiz Villareal

3.4 The Cell Membrane

Learning Objectives

By the end of this section, you will be able to:

- Understand the fluid mosaic model of membranes
- Describe the functions of phospholipids, proteins, and carbohydrates in membranes

A cell's plasma membrane defines the boundary of the cell and determines the nature of its contact with the environment. Cells exclude some substances, take in others, and excrete still others, all in controlled quantities. Plasma membranes enclose the borders of cells, but rather than being a static bag, they are dynamic and constantly in flux. The plasma membrane must be sufficiently 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 as tissues and organs form during early development, and which later plays a role in the "self" versus "non-self" distinction of the immune response.

The plasma membrane also carries receptors, which are attachment sites for specific substances that interact with the cell. Each receptor is structured to bind with a specific substance. For example, surface receptors of the membrane create changes in the interior, such as changes in enzymes of metabolic pathways. These metabolic pathways might be vital for providing the cell with energy, making specific substances for the cell, or breaking down cellular waste or toxins for disposal. Receptors on the plasma membrane's exterior surface interact with hormones or neurotransmitters, and allow their messages to be transmitted into the cell. Some recognition sites are used by viruses as attachment points. Although they are highly specific, pathogens like viruses may evolve to exploit receptors to gain entry to a cell by mimicking the specific substance that the receptor is meant to bind. This specificity helps to explain why human immunodeficiency virus (HIV) or any of the five types of hepatitis viruses invade only specific cells.

Fluid Mosaic Model

In 1972, S. J. Singer and Garth L. Nicolson proposed a new model of the plasma membrane that, compared to earlier understanding, better explained both microscopic observations and the function of the plasma membrane. This was called the fluid mosaic model. The model has evolved somewhat over time, but 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—in which the components are able to flow and change position, while maintaining the basic integrity of the membrane. Both phospholipid molecules and embedded proteins are able to diffuse rapidly and laterally in the membrane. The fluidity of the plasma membrane is necessary for the activities of certain enzymes and transport molecules within the membrane. Plasma membranes range from 5–10 nm thick. As a comparison, human red blood cells, visible via light microscopy, are approximately 8 µm thick, or approximately 1,000 times thicker than a plasma membrane.

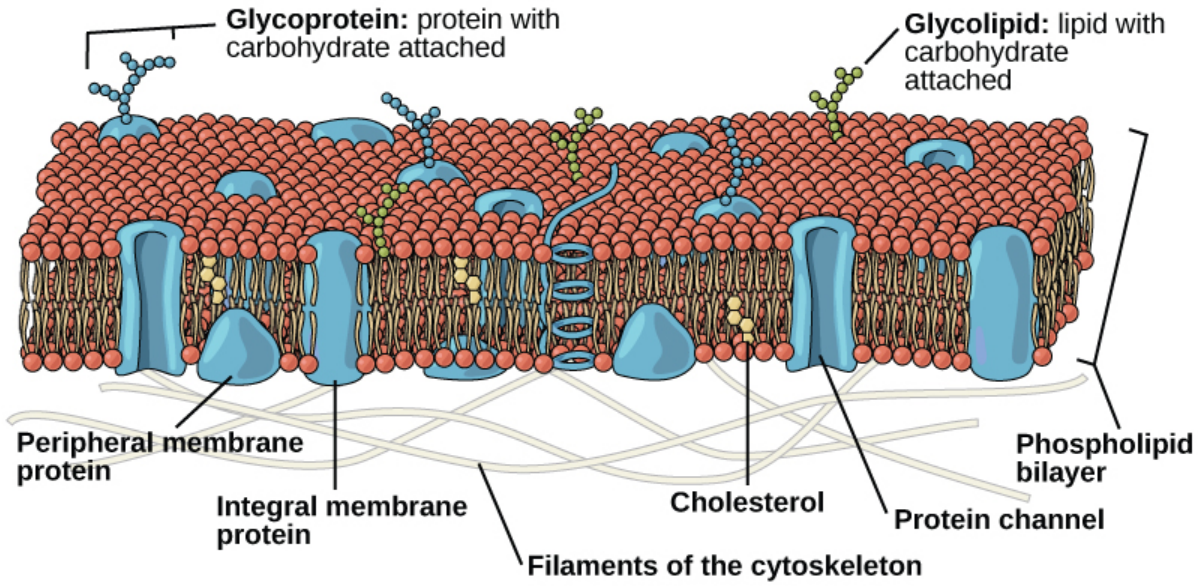


Figure 3.21 The fluid mosaic model of the plasma membrane structure describes the plasma membrane as a fluid combination of phospholipids, cholesterol, proteins, and carbohydrates.

The plasma membrane is made up primarily of a bilayer of phospholipids with embedded proteins, carbohydrates, glycolipids, and glycoproteins, and, in animal cells, cholesterol. The amount of cholesterol in animal plasma membranes regulates the fluidity of the membrane and changes based on the temperature of the cell's environment. In other words, cholesterol acts as antifreeze in the cell membrane and is more abundant in animals that live in cold climates.

The main fabric of the membrane is composed of two layers of phospholipid molecules, and the polar ends of these molecules (which look like a collection of balls in an artist's rendition of the model) (Figure 3.22) are in contact with aqueous fluid both inside and outside the cell. Thus, both surfaces of the plasma membrane are hydrophilic. In contrast, the interior of the membrane, between its two surfaces, is a hydrophobic or nonpolar region because of the fatty acid tails. This region has no attraction for water or other polar molecules.

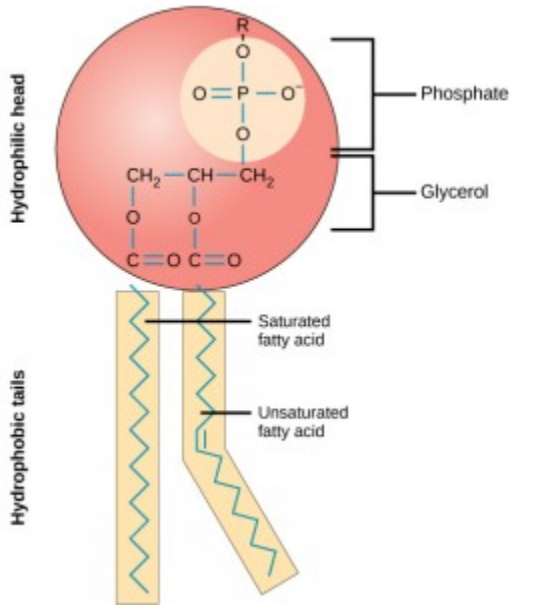


Figure 3.22 This phospholipid molecule is composed of a hydrophilic head and two hydrophobic tails. The hydrophilic head group consists of a phosphate-containing group attached to a glycerol molecule. The hydrophobic tails, each containing either a saturated or an unsaturated fatty acid, are long hydrocarbon chains.

Proteins make up the second major chemical component of plasma membranes. Integral proteins are embedded in the plasma membrane and may span all or part of the membrane. Integral proteins may serve as channels or pumps to move materials into or out of the cell. Peripheral proteins are found on the exterior or interior surfaces of membranes, attached either to integral proteins or to phospholipid molecules. Both integral and peripheral proteins may serve as enzymes, as structural attachments for the fibers of the cytoskeleton, or as part of the cell's recognition sites.

Carbohydrates are the third major component of plasma membranes. They are always found on the exterior surface of cells and are bound either to proteins (forming glycoproteins) or to lipids (forming glycolipids). These carbohydrate chains may consist of 2–60 monosaccharide units and may be either straight or branched. Along with peripheral proteins, carbohydrates form specialized sites on the cell surface that allow cells to recognize each other.

Evolution in Action

How Viruses Infect Specific OrgansSpecific glycoprotein molecules exposed on the surface of the cell membranes of host cells are exploited by many viruses to infect specific organs. For example, HIV is able to penetrate the plasma membranes of specific kinds of white blood cells called T-helper cells and monocytes, as well as some cells of the central nervous system. The hepatitis virus attacks only liver cells.

These viruses are able to invade these cells, because the cells have binding sites on their surfaces that

the viruses have exploited with equally specific glycoproteins in their coats. (Figure 3.23). The cell is tricked by the mimicry of the virus coat molecules, and the virus is able to enter the cell. Other recognition sites on the virus's surface interact with the human immune system, prompting the body to produce antibodies. Antibodies are made in response to the antigens (or proteins associated with invasive pathogens). These same sites serve as places for antibodies to attach, and either destroy or inhibit the activity of the virus. Unfortunately, these sites on HIV are encoded by genes that change quickly, making the production of an effective vaccine against the virus very difficult. The virus population within an infected individual quickly evolves through mutation into different populations, or variants, distinguished by differences in these recognition sites. This rapid change of viral surface markers decreases the effectiveness of the person's immune system in attacking the virus, because the antibodies will not recognize the new variations of the surface patterns.

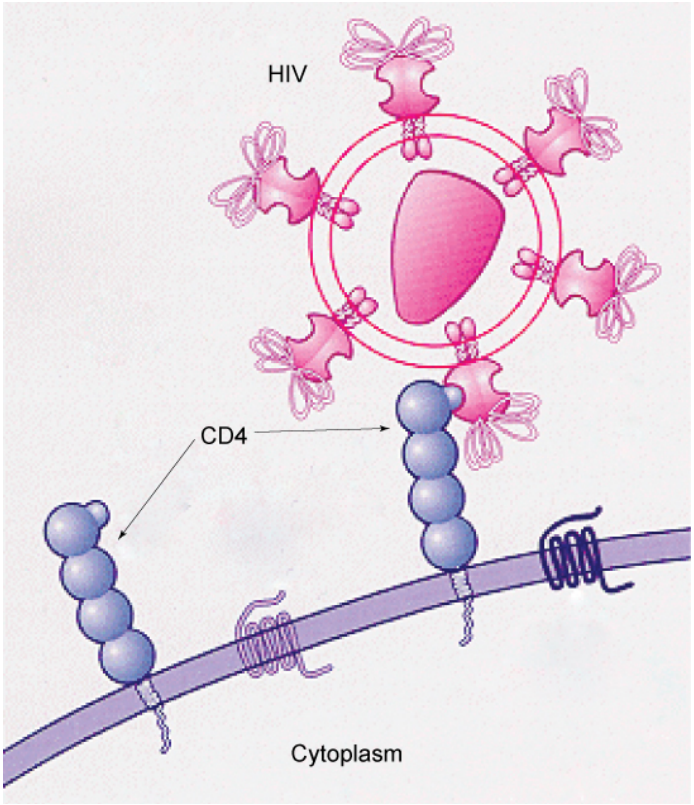


Figure 3.23 HIV docks at and binds to the CD4 receptor, a glycoprotein on the surface of T cells, before entering, or infecting, the cell.

Section Summary

The modern understanding of the plasma membrane is referred to as the fluid mosaic model. The plasma membrane is composed of a bilayer of phospholipids, with their hydrophobic, fatty acid tails in contact with each other. The landscape of the membrane is studded with proteins, some of which span the membrane. Some of these proteins serve to transport materials into or out of the cell. Carbohydrates are attached to some of the proteins and lipids on the outward-facing surface of the membrane. These form complexes that function to identify the cell to other cells. The fluid nature of the membrane owes itself to the configuration of the fatty acid tails, the presence of cholesterol embedded in the membrane

(in animal cells), and the mosaic nature of the proteins and protein-carbohydrate complexes, which are not firmly fixed in place. Plasma membranes enclose the borders of cells, but rather than being a static bag, they are dynamic and constantly in flux.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4402#h5p-23

Glossary

fluid mosaic model: a model of the structure of the plasma membrane as a mosaic of components, including phospholipids, cholesterol, proteins, and glycolipids, resulting in a fluid rather than static character

Media Attribution

- Figure 3.23: modification of work by US National Institutes of Health/National Institute of Allergy and Infectious Diseases

3.5 Passive Transport

Learning Objectives

By the end of this section, you will be able to:

- Explain why and how passive transport occurs
- Understand the processes of osmosis and diffusion
- Define tonicity and describe its relevance to passive transport

Plasma membranes must allow certain substances to enter and leave a cell, while preventing harmful material from entering and essential material from leaving. In other words, plasma membranes are selectively permeable—they allow some substances 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 the extracellular fluids. This may happen passively, as certain materials move back and forth, or the cell may have special mechanisms that ensure transport. Most cells expend most of their energy, in the form of adenosine triphosphate (ATP), to create and maintain an uneven distribution of ions on the opposite sides of their membranes. The structure of the plasma membrane contributes to these functions, but it also presents some problems.

The most direct forms of membrane transport are passive. **Passive transport is a naturally occurring phenomenon and does not require the cell to expend energy** to accomplish the movement. In passive transport, substances move from an area of higher concentration to an area of lower concentration in a process called **diffusion**. A physical space in which there is a different concentration of a single substance is said to have a concentration gradient.

Selective Permeability

Plasma membranes are asymmetric, meaning that despite the mirror image formed by the phospholipids, the interior of the membrane is not identical to the exterior of the membrane. Integral proteins that act as channels or pumps work in one direction. 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.

Recall that plasma membranes have hydrophilic and hydrophobic regions. This characteristic helps the movement of certain materials through the membrane and hinders the movement of others. **Lipid-**

soluble material 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 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 pass through by simple diffusion.

Polar substances, with the exception of water, 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, whereas 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 a special means of penetrating plasma membranes. Simple sugars and amino acids also need help with transport across plasma membranes.

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 the space. You are familiar with diffusion of substances through the air. For example, think about someone opening a bottle of perfume in a room filled with people. The perfume is at its highest concentration in the bottle and is at its lowest at the edges of the room. The perfume vapor will diffuse, or spread away, from the bottle, and gradually, more and more people will smell the perfume as it spreads. Materials move within the cell's cytosol by diffusion, and certain materials move through the plasma membrane by diffusion (Figure 3.24). Diffusion expends no energy. Rather the different concentrations of materials in different areas are a form of potential energy, and diffusion is the dissipation of that potential energy as materials move down their concentration gradients, from high to low.

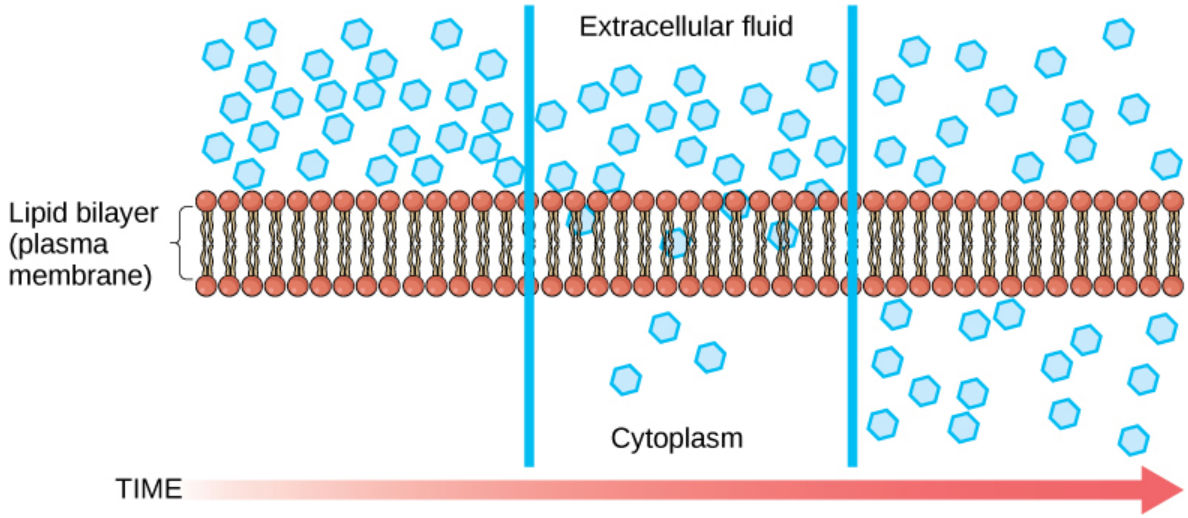


Figure 3.24 Diffusion through a permeable membrane follows the concentration gradient of a substance, moving the substance from an area of high concentration to one of low concentration.

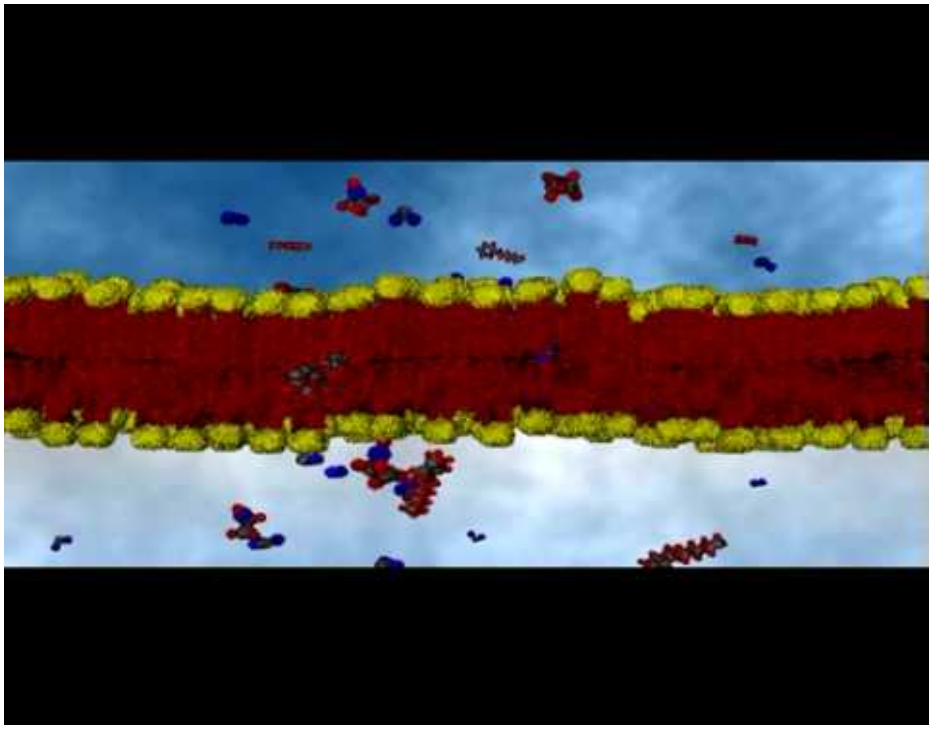
Each separate substance in a medium, such as the extracellular fluid, has its own concentration gradient, independent of the concentration gradients of other materials. Additionally, each substance will diffuse according to that gradient.

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: More massive molecules move more slowly, because it is more difficult for them to move between the molecules of the substance they are moving through; therefore, they diffuse more slowly.
- Temperature: Higher temperatures increase the energy and therefore the movement of the molecules, increasing the rate of diffusion.
- Solvent density: As the density of the solvent increases, the rate of diffusion decreases. The molecules slow down because they have a more difficult time getting through the denser medium.

Concept in Action

For an animation of the diffusion process in action, view this short video on cell membrane transport.



A YouTube element has been excluded from this version of the text. You can view it online here: https://opentextbc.ca/biology/?p=4409

Facilitated transport

In facilitated transport, also called facilitated diffusion, material moves across the plasma membrane **with the assistance of transmembrane proteins** down a concentration gradient (from high to low concentration) without the expenditure of cellular energy. However, the substances that undergo facilitated transport would otherwise not diffuse easily or quickly across the plasma membrane. The solution to moving polar substances and other substances across the plasma membrane rests in the proteins that span its surface. 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, because they form channels or pores that allow certain substances to pass through the membrane. 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.

Osmosis

Osmosis is the **diffusion of water** through a semipermeable membrane according to the concentration gradient of water across the membrane. Whereas 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. Osmosis is a special case of diffusion. Water, like other substances, moves from an area of higher concentration to one of lower concentration. Imagine a beaker with a semipermeable membrane, separating the two sides or halves (Figure 3.25). On both sides of the membrane, the water level is the same, but there are different concentrations on each side of a dissolved substance, or solute, that cannot cross the membrane. If the volume of the water is the same, but the concentrations of solute are different, then there are also different concentrations of water, the solvent, on either side of the membrane.

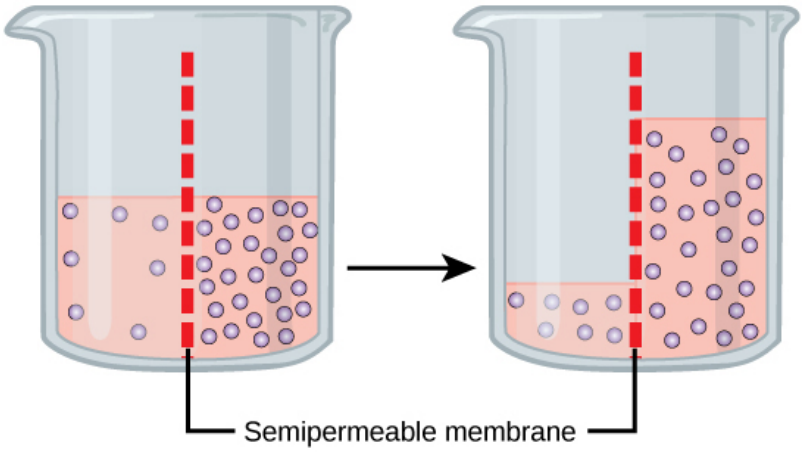


Figure 3.25 In osmosis, water always moves from an area of higher concentration (of water) to one of lower concentration (of water). In this system, the solute cannot pass through the selectively permeable membrane.

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. Therefore, 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. Osmosis proceeds constantly in living systems.

Tonicity

Tonicity describes the amount of solute in a solution. The measure of the tonicity of a solution, or the total amount of solutes dissolved in a specific amount of solution, is called its osmolarity. 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 solution, such as tap water, the extracellular fluid has a lower concentration of solutes 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 than does the cell. In this situation, water will follow its concentration gradient and enter the cell. This may cause an animal cell to burst, or lyse.

In a hypertonic solution (the prefix *hyper–* refers to the extracellular fluid having a higher concentration of solutes than the cell's cytoplasm), the fluid contains less water than the cell does, such as seawater. Because the cell has a lower concentration of solutes, the water will leave the cell. In effect, the solute is drawing the water out of the cell. This may cause an animal cell to shrivel, or crenate.

In an isotonic solution, the extracellular fluid has the same osmolarity as the cell. If the concentration of solutes of the cell matches that of the extracellular fluid, there will be no net movement of water into or out of the cell. Blood cells in hypertonic, isotonic, and hypotonic solutions take on characteristic appearances (Figure 3.26).

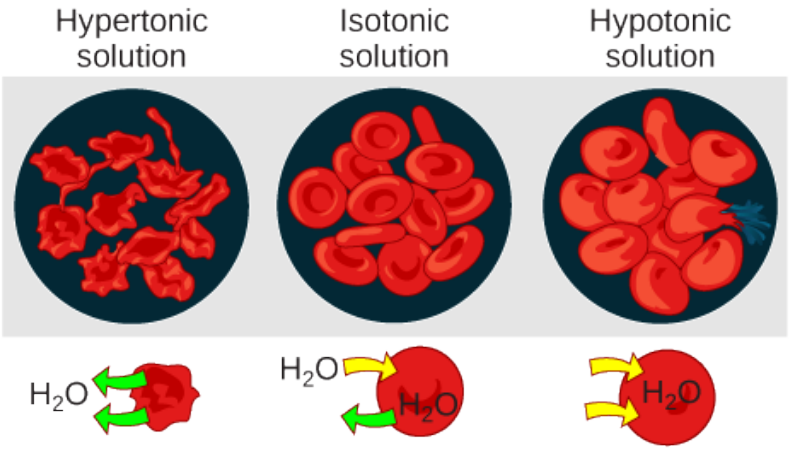


Figure 3.26 Osmotic pressure changes the shape of red blood cells in hypertonic, isotonic, and hypotonic solutions.

A doctor injects a patient with what the doctor thinks is isotonic saline solution. The patient dies, and autopsy reveals that many red blood cells have been destroyed. Do you think the solution the doctor injected was really isotonic?

<!– No, it must have been hypotonic, as a hypotonic solution would cause water to enter the cells, thereby making them burst. –>

Some organisms, such as plants, fungi, bacteria, and some protists, have cell walls that surround the plasma membrane and prevent cell lysis. 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 compared to the cellular environment, and water will always enter a cell if water is available. This influx of water produces turgor pressure, which stiffens the cell walls of the plant (Figure 3.27). In nonwoody plants, turgor pressure supports the plant. If the plant cells become hypertonic, as occurs in drought or if a plant is not watered adequately, water will leave the cell. Plants lose turgor pressure in this condition and wilt.

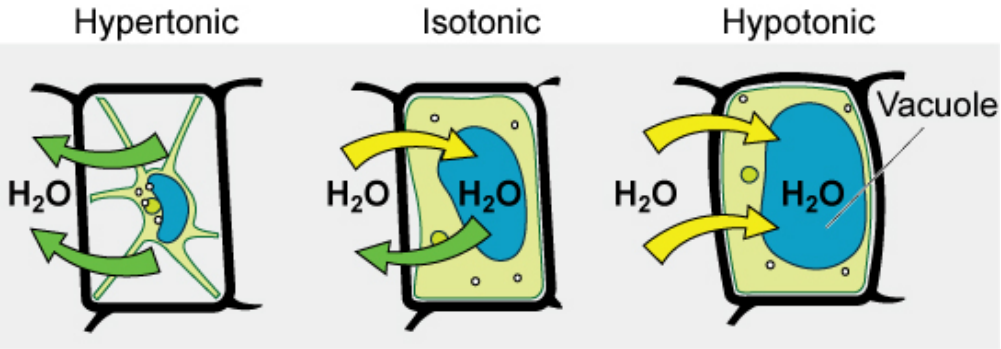


Figure 3.27 The turgor pressure within a plant cell depends on the tonicity of the solution that it is bathed in.

Section Summary

The passive forms of transport, diffusion and osmosis, move material of small molecular weight. Substances diffuse from areas of high concentration to areas of low concentration, and this process continues until the substance is evenly distributed in a system. In solutions of more than one substance, each type of molecule diffuses according to its own concentration gradient. Many factors can affect the rate of diffusion, including concentration gradient, the sizes of the particles that are diffusing, and the temperature of the system.

In living systems, diffusion of substances into and out of cells is mediated by the plasma membrane. Some materials diffuse readily through the membrane, but others are hindered, and their passage is only made possible by protein channels and carriers. The chemistry of living things occurs in aqueous solutions, and balancing the concentrations of those solutions is an ongoing problem. In living systems, diffusion of some substances would be slow or difficult without membrane proteins.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4409#h5p-24

Glossary

concentration gradient: an area of high concentration across from an area of low concentration

diffusion: a passive process of transport of low-molecular weight material down its concentration gradient

facilitated transport: a process by which material moves down a concentration gradient (from high to low concentration) using integral membrane proteins

hypertonic: describes a solution in which extracellular fluid has higher osmolarity than the fluid inside the cell

hypotonic: describes a solution in which extracellular fluid has lower osmolarity than the fluid inside the cell

isotonic: describes a solution in which the extracellular fluid has the same osmolarity as the fluid inside the cell

osmolarity: the total amount of substances dissolved in a specific amount of solution

osmosis: the transport of water through a semipermeable membrane from an area of high water concentration to an area of low water concentration across a membrane

passive transport: a method of transporting material that does not require energy

selectively permeable: the characteristic of a membrane that allows some substances through but not others

solute: a substance dissolved in another to form a solution

tonicity: the amount of solute in a solution.

Media Attributions

- Figure 3.24: modification of work by Mariana Ruiz Villarreal
- Figure 3.26: modification of work by Mariana Ruiz Villarreal
- Figure 3.27: modification of work by Mariana Ruiz Villarreal

3.6 Active Transport

Learning Objectives

By the end of this section, you will be able to:

- Understand how electrochemical gradients affect ions
- Describe endocytosis, including phagocytosis, pinocytosis, and receptor-mediated endocytosis
- Understand the process of exocytosis

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 3.28). The combined gradient that affects an ion is called its electrochemical gradient, and it is especially important to muscle and nerve cells.

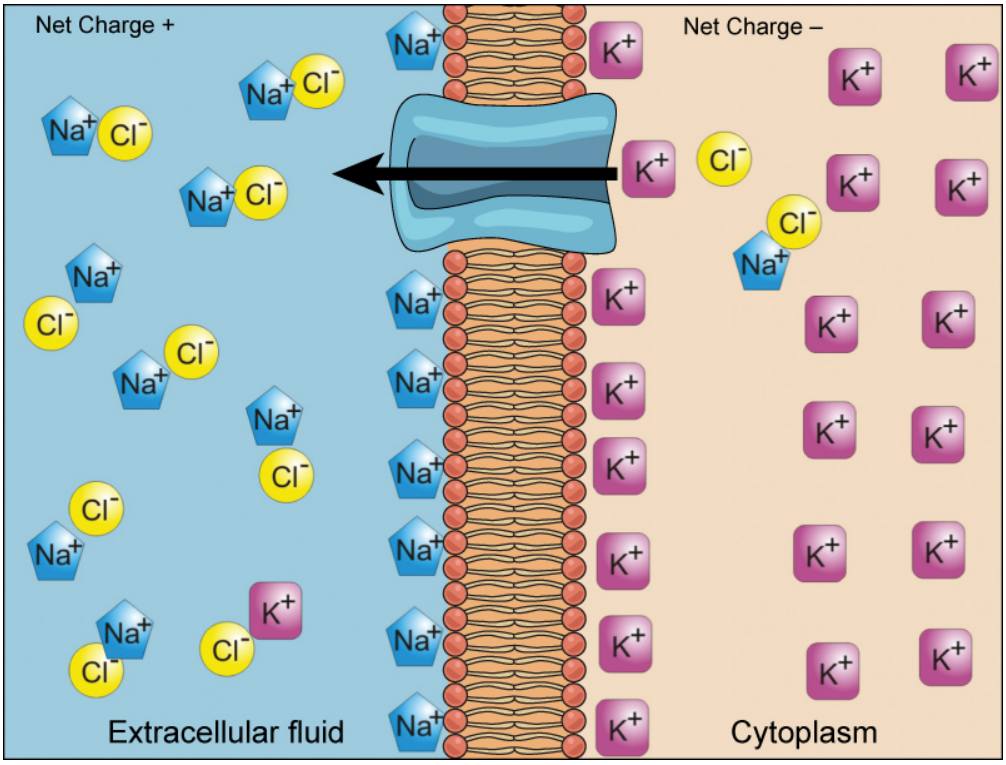


Figure 3.28 Electrochemical gradients arise from the combined effects of concentration gradients and electrical gradients.

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 3.29). The action of this pump results in a concentration and charge difference across the membrane.

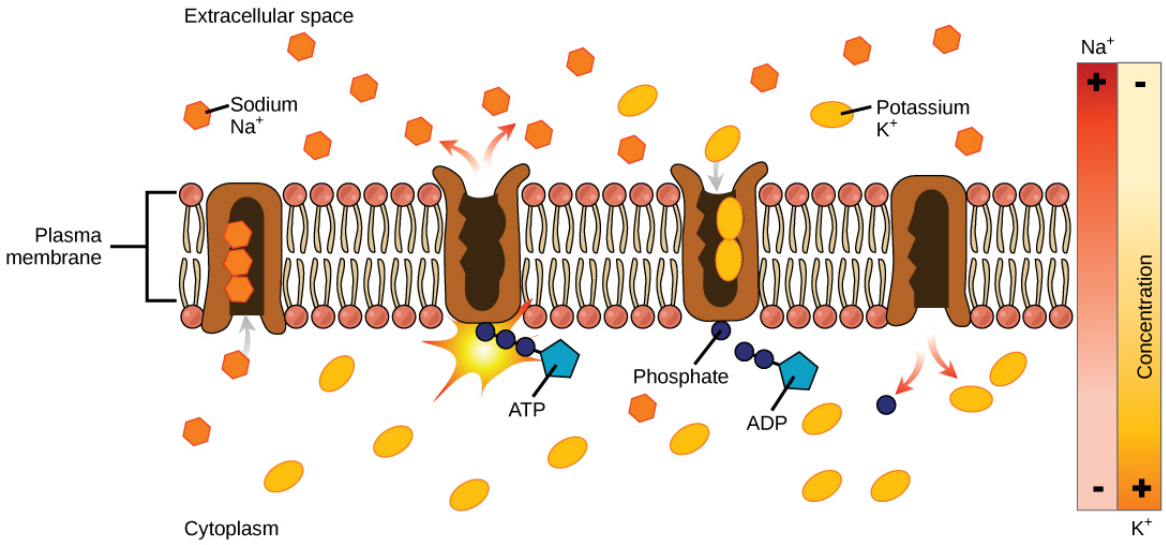


Figure 3.29 The sodium-potassium pump move potassium and sodium ions across the plasma membrane.

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.

Endocytosis

Endocytosis is a type of active 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.

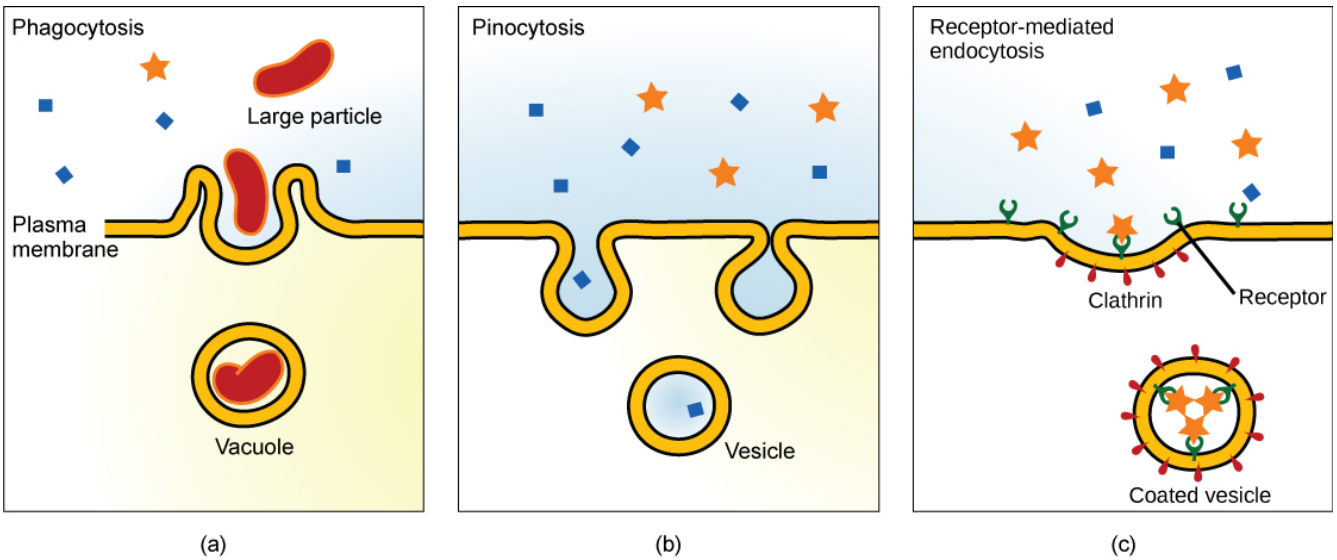


Figure 3.30 Three variations of endocytosis are shown. (a) In one form of endocytosis, phagocytosis, the cell membrane surrounds the particle and pinches off to form an intracellular vacuole. (b) In another type of endocytosis, pinocytosis, the cell membrane surrounds a small volume of fluid and pinches 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 external cell membrane.

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 3.30).

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 3.30).

A targeted variation of endocytosis employs binding proteins in the plasma membrane that are specific for certain substances (Figure 3.30). 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 3.31).

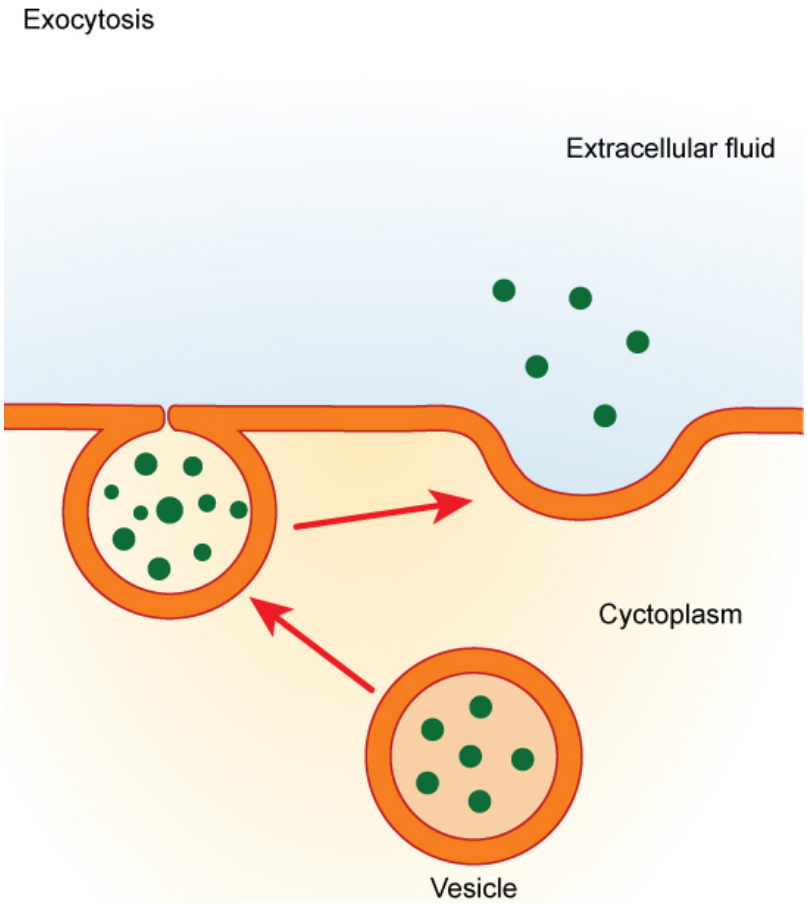


Figure 3.31 In exocytosis, a vesicle migrates to the plasma membrane, binds, and releases its contents to the outside of the cell.

Section Summary

The combined gradient that affects an ion includes its concentration gradient and its electrical gradient. Living cells need certain substances in concentrations greater than they exist in the extracellular space. Moving substances up their electrochemical gradients requires energy from the cell. Active transport uses energy stored in ATP to fuel the transport. Active transport of small molecular-size material uses integral proteins in the cell membrane to move the material—these proteins are analogous to pumps. Some pumps, which carry out primary active transport, couple directly with ATP to drive their action. In secondary transport, energy from primary transport can be used to move another substance into the cell and up its concentration gradient.

Endocytosis methods require the direct use of ATP to fuel the transport of large particles such as macromolecules; parts of cells or whole cells can be engulfed by other cells in a process called phagocytosis. In phagocytosis, a portion of the membrane invaginates and flows around the particle,

eventually pinching off and leaving the particle wholly enclosed by an envelope of plasma membrane. Vacuoles are broken down by the cell, with the particles used as food or dispatched in some other way. Pinocytosis is a similar process on a smaller scale. The cell expels waste and other particles through the reverse process, exocytosis. Wastes are moved outside the cell, pushing a membranous vesicle to the plasma membrane, allowing the vesicle to fuse with the membrane and incorporating itself into the membrane structure, releasing its contents to the exterior of the cell.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4415#h5p-25

Glossary

active transport: the method of transporting material that requires energy

electrochemical gradient: a gradient produced by the combined forces of the electrical gradient and the chemical gradient

endocytosis: a type of active transport that moves substances, including fluids and particles, into a cell

exocytosis: a process of passing material out of a cell

phagocytosis: a process that takes macromolecules that the cell needs from the extracellular fluid; a variation of endocytosis

pinocytosis: a process that takes solutes that the cell needs from the extracellular fluid; a variation of endocytosis

receptor-mediated endocytosis: a variant of endocytosis that involves the use of specific binding proteins in the plasma membrane for specific molecules or particles

Media Attributions

- Figure 3.28: modification of work by “Synaptitude”/Wikimedia Commons
- Figure 3.29: modification of work by Mariana Ruiz Villarreal
- Figure 3.30: modification of work by Mariana Ruiz Villarreal
- Figure 3.31: modification of work by Mariana Ruiz Villarreal

Chapter 4: Introduction to How Cells Obtain Energy

Figure 4.1 A hummingbird needs energy to maintain prolonged flight. The bird obtains its energy from taking in food and transforming the energy contained in food molecules into forms of energy to power its flight through a series of biochemical reactions. (credit: modification of work by Cory Zanker)

Virtually every task performed by living organisms requires energy. Energy is needed to perform heavy labor and exercise, but humans also use energy while thinking, and even during sleep. In fact, the living cells of every organism constantly use energy. Nutrients and other molecules are imported into the cell, metabolized (broken down) and possibly synthesized into new molecules, modified if needed, transported around the cell, and possibly distributed to the entire organism. For example, the large proteins that make up muscles are built from smaller molecules imported from dietary amino acids. Complex carbohydrates are broken down into simple sugars that the cell uses for energy. Just as energy is required to both build and demolish a building, energy is required for the synthesis and breakdown of molecules as well as the transport of molecules into and out of cells. In addition, processes such as ingesting and breaking down pathogenic bacteria and viruses, exporting wastes and toxins, and movement of the cell require energy. From where, and in what form, does this energy come? How do living cells obtain energy, and how do they use it? This chapter will discuss different forms of energy and the physical laws that govern energy transfer. This chapter will also describe how cells use energy and replenish it, and how chemical reactions in the cell are performed with great efficiency.

Search for Key Points in Chapter 4

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4418#h5p-26

4.1 Energy and Metabolism

Learning Objectives

By the end of this section, you will be able to:

- Explain what metabolic pathways are
- State the first and second laws of thermodynamics
- Explain the difference between kinetic and potential energy
- Describe endergonic and exergonic reactions
- Discuss how enzymes function as molecular catalysts

Watch a video about heterotrophs.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4435#h5p-27

Scientists use the term bioenergetics to describe the concept of energy flow (Figure 4.2) through living systems, such as cells. **Cellular processes** such as the building and breaking down of complex molecules **occur through stepwise chemical reactions**. Some of these chemical reactions are spontaneous and release energy, whereas others require energy to proceed. Just as living things must continually consume food to replenish their energy supplies, cells must continually produce more energy to replenish that used by the many energy-requiring chemical reactions that constantly take place. Together, **all of the chemical reactions** that take place inside cells, including those that consume or generate energy, are referred to as the **cell's metabolism**.

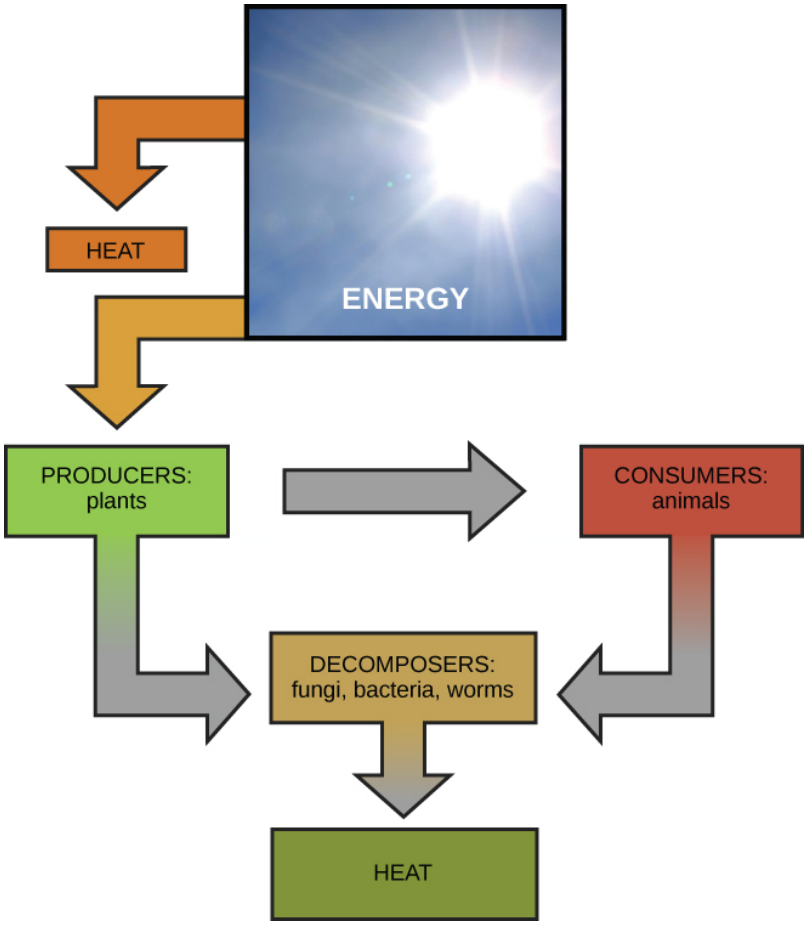


Figure 4.2 Ultimately, most life forms get their energy from the sun. Plants use photosynthesis to capture sunlight, and herbivores eat the plants to obtain energy. Carnivores eat the herbivores, and eventual decomposition of plant and animal material contributes to the nutrient pool.

Metabolic Pathways

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

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

Because this process involves synthesizing an energy-storing molecule, it requires energy input to proceed. During the light reactions of photosynthesis, **energy is provided by a molecule called adenosine triphosphate (ATP)**, which is the primary energy currency of all cells. Just as the dollar is used as currency to buy goods, cells use molecules of ATP as energy currency to perform immediate work. In contrast, energy-storage molecules such as glucose are consumed only to be broken down to

use their energy. The reaction that harvests the energy of a sugar molecule in cells requiring oxygen to survive can be summarized by the reverse reaction to photosynthesis. In this reaction, oxygen is consumed and carbon dioxide is released as a waste product. The reaction is summarized as:

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

Both of these reactions involve many steps.

The processes of making and breaking down sugar molecules illustrate two examples of metabolic pathways. A metabolic pathway is a series of chemical reactions that takes a starting molecule and modifies it, step-by-step, through a series of metabolic intermediates, eventually yielding a final product. In the example of sugar metabolism, the first metabolic pathway synthesized sugar from smaller molecules, and the other pathway broke sugar down into smaller molecules. These two opposite processes—the first requiring energy and the second producing energy—are referred to as **anabolic pathways (building polymers) and catabolic pathways (breaking down polymers into their monomers)**, respectively. Consequently, metabolism is composed of synthesis (anabolism) and degradation (catabolism) (Figure 4.3).

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

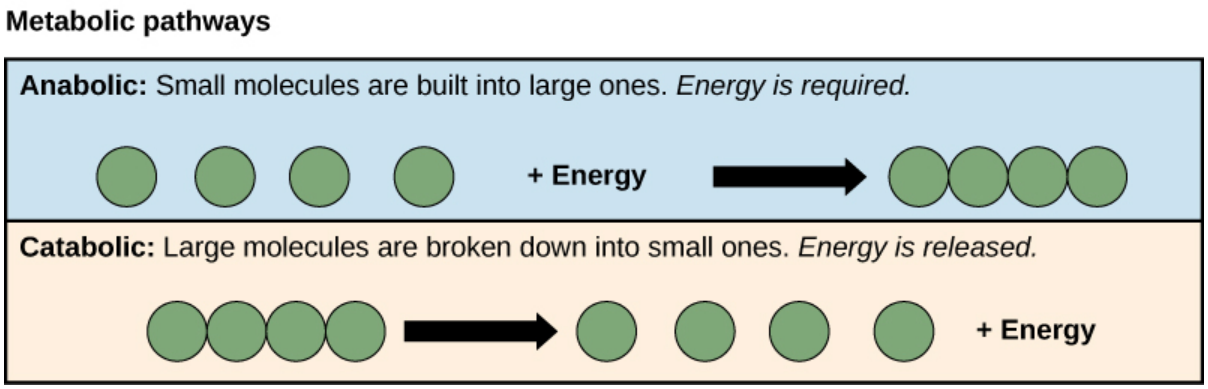


Figure 4.3 Catabolic pathways are those that generate energy by breaking down larger molecules. Anabolic pathways are those that require energy to synthesize larger molecules. Both types of pathways are required for maintaining the cell's energy balance.

Energy

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

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

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

Thermodynamics

The first law of thermodynamics states that the total amount of energy in the universe is constant and conserved. In other words, there has always been, and always will be, exactly the same amount of energy in the universe. **Energy exists in many different forms**. According to the first law of thermodynamics, energy may be transferred from place to place or transformed into different forms, **but it cannot be created or destroyed**. The transfers and transformations of energy take place around us all the time. Light bulbs transform electrical energy into light and heat energy. Gas stoves transform chemical energy from natural gas into heat energy. Plants perform one of the most biologically useful energy transformations on earth: that of converting the energy of sunlight to chemical energy stored within organic molecules (Figure 4.2). Some examples of energy transformations are shown in Figure 4.4.

The challenge for all living organisms is to obtain energy from their surroundings in forms that they can transfer or transform into usable energy to do work. Living cells have evolved to meet this challenge. Chemical energy stored within organic molecules such as sugars and fats is transferred and transformed through a series of cellular chemical reactions into energy within molecules of ATP. Energy in ATP molecules is easily accessible to do work. Examples of the types of work that cells need to do include building complex molecules, transporting materials, powering the motion of cilia or flagella, and contracting muscle fibers to create movement.

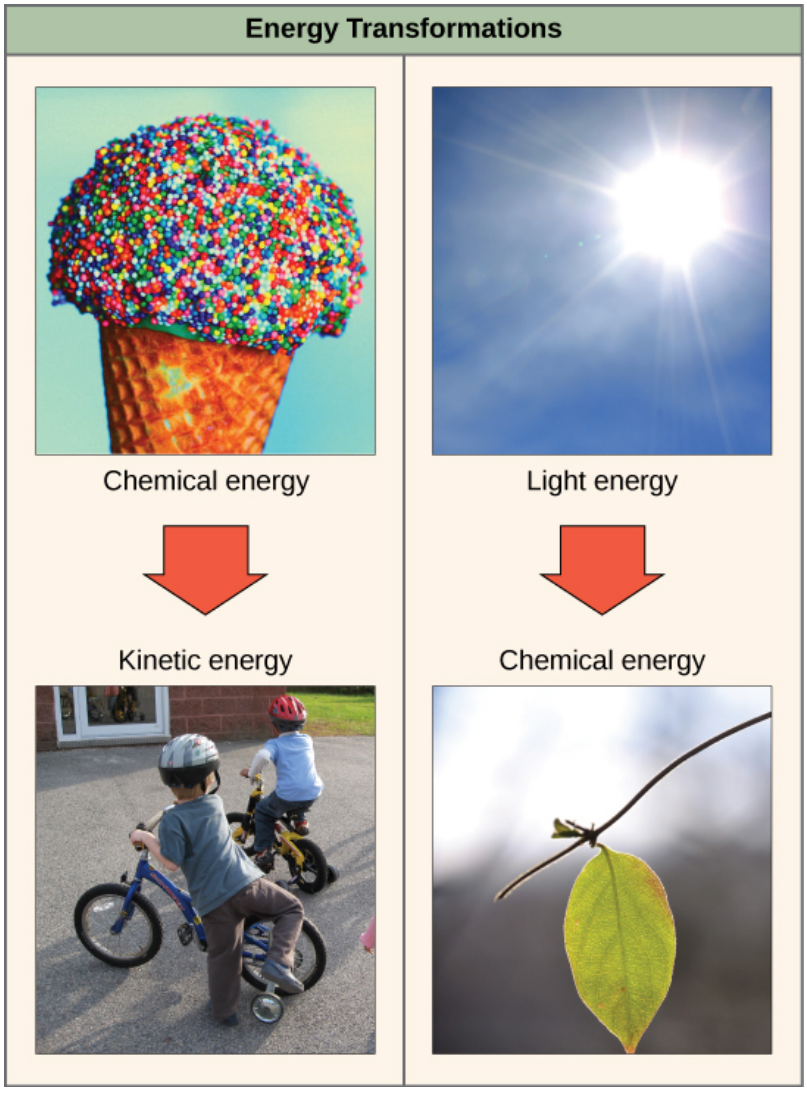


Figure 4.4 Shown are some examples of energy transferred and transformed from one system to another and from one form to another. The food we consume provides our cells with the energy required to carry out bodily functions, just as light energy provides plants with the means to create the chemical energy they need. (credit "ice cream": modification of work by D. Sharon Pruitt; credit "kids": modification of work by Max from Providence; credit "leaf": modification of work by Cory Zanker)

A living cell's primary tasks of obtaining, transforming, and using energy to do work may seem simple. However, the **second law of thermodynamics** explains why these tasks are harder than they appear. **All energy transfers and transformations are never completely efficient**. In every energy transfer, some amount of energy is lost in a form that is unusable. In most cases, this form is heat energy. Thermodynamically, heat energy is defined as the energy transferred from one system to another that is not work. For example, when a light bulb is turned on, some of the energy being converted from electrical energy into light energy is lost as heat energy. Likewise, some energy is lost as heat energy during cellular metabolic reactions.

An important concept in physical systems is that of order and disorder. The more energy that is lost by

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

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

Potential and Kinetic Energy

When an object is in motion, there is energy associated with that object. Think of a wrecking ball. Even a slow-moving wrecking ball can do a great deal of damage to other objects. Energy associated with objects in motion is called kinetic energy (Figure 4.5). A speeding bullet, a walking person, and the rapid movement of molecules in the air (which produces heat) all have kinetic energy.

Now what if that same motionless wrecking ball is lifted two stories above ground with a crane? If the suspended wrecking ball is unmoving, is there energy associated with it? The answer is yes. The energy that was required to lift the wrecking ball did not disappear, but is now stored in the wrecking ball by virtue of its position and the force of gravity acting on it. This type of energy is called potential energy (Figure 4.5). If the ball were to fall, the potential energy would be transformed into kinetic energy until all of the potential energy was exhausted when the ball rested on the ground. Wrecking balls also swing like a pendulum; through the swing, there is a constant change of potential energy (highest at the top of the swing) to kinetic energy (highest at the bottom of the swing). Other examples of potential energy include the energy of water held behind a dam or a person about to skydive out of an airplane.

Figure 4.5 Still water has potential energy; moving water, such as in a waterfall or a rapidly flowing river, has kinetic energy. (credit "dam": modification of work by "Pascal"/Flickr; credit "waterfall": modification of work by Frank Gualtieri)

Potential energy is not only associated with the location of matter, but also with the structure of matter. Even a spring on the ground has potential energy if it is compressed; so does a rubber band that is

pulled taut. On a molecular level, the bonds that hold the atoms of molecules together exist in a particular structure that has potential energy. Remember that anabolic cellular pathways require energy to synthesize complex molecules from simpler ones and catabolic pathways release energy when complex molecules are broken down. The fact that energy can be released by the breakdown of certain chemical bonds implies that those bonds have potential energy. In fact, there is potential energy stored within the bonds of all the food molecules we eat, which is eventually harnessed for use. This is because these bonds can release energy when broken. The type of potential energy that exists within chemical bonds, and is released when those bonds are broken, is called chemical energy. Chemical energy is responsible for providing living cells with energy from food. The release of energy occurs when the molecular bonds within food molecules are broken.

Watch a video about kilocalories.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4435#h5p-28

Concept in Action

Visit the site and select “Pendulum” from the “Work and Energy” menu to see the shifting kinetic and potential energy of a pendulum in motion.

Free and Activation Energy

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

If energy is released during a chemical reaction, then the change in free energy, signified as ΔG (delta G) will be a negative number. A negative change in free energy also means that the products of the

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

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

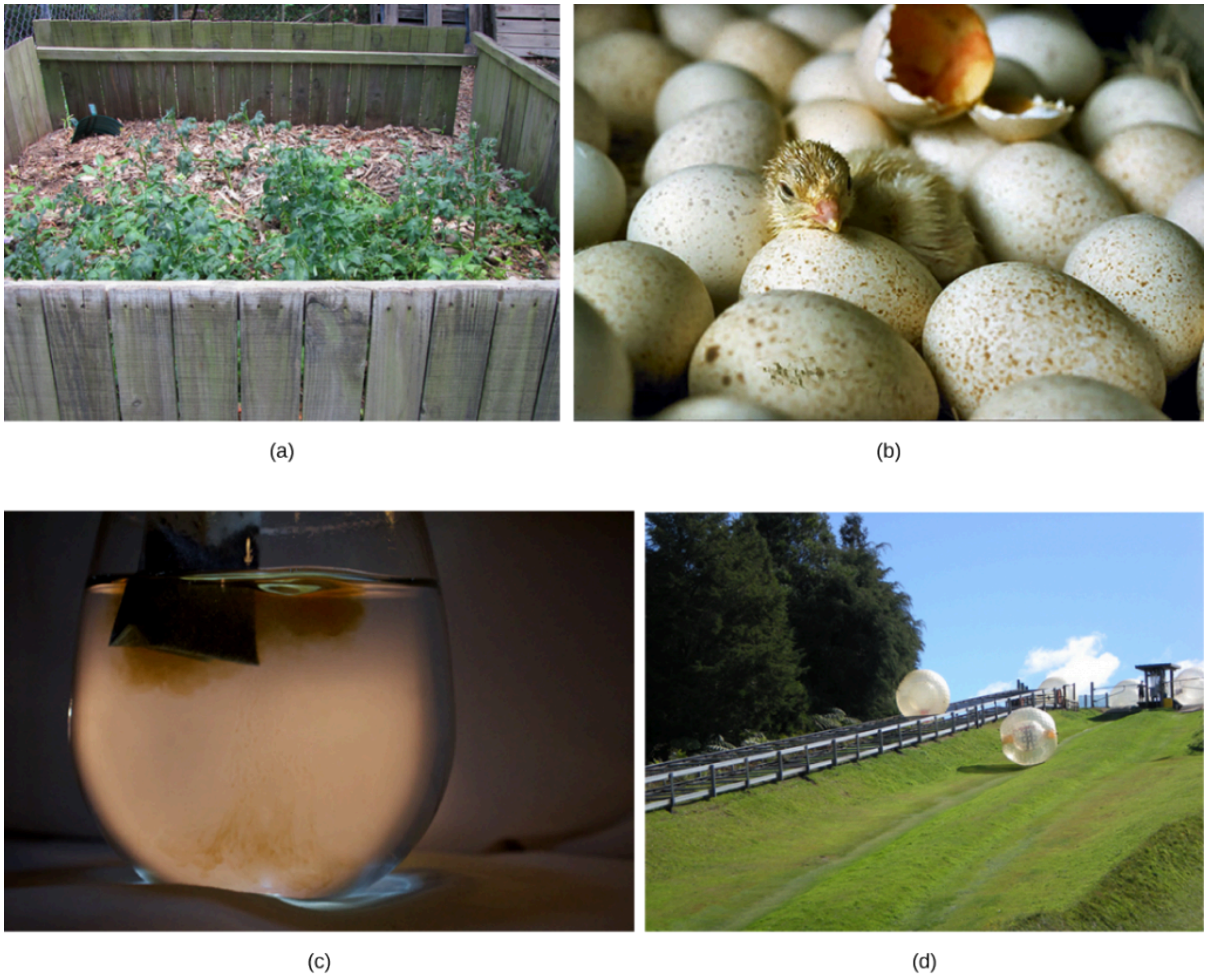


Figure 4.6 Shown are some examples of endergonic processes (ones that require energy) and exergonic processes (ones that release energy). (credit a: modification of work by Natalie Maynor; credit b: modification of work by USDA; credit c: modification of work by Cory Zanker; credit d: modification of work by Harry Malsch)

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

There is another important concept that must be considered regarding endergonic and exergonic reactions. Exergonic reactions require a small amount of energy input to get going, before they can proceed with their energy-releasing steps. These reactions have a net release of energy, but still require some energy input in the beginning. This small amount of energy input necessary for all chemical reactions to occur is called the activation energy.

Concept in Action

Watch an animation of the move from free energy to transition state of the reaction.

Enzymes

A substance that helps a chemical reaction to occur is called a catalyst, and the molecules that catalyze biochemical reactions are called enzymes. Most enzymes are **proteins** and perform the critical task of **lowering the activation energies** of chemical reactions inside the cell. Most of the reactions critical to a living cell happen too slowly at normal temperatures to be of any use to the cell. Without enzymes to **speed up these reactions**, life could not persist. Enzymes do this by binding to the reactant molecules and holding them in such a way as to make the chemical bond-breaking and -forming processes take place more easily. It is important to remember that enzymes do not change whether a reaction is exergonic (spontaneous) or endergonic. This is because they do not change the free energy of the reactants or products. They only reduce the activation energy required for the reaction to go forward (Figure 4.7). In addition, an enzyme itself is unchanged by the reaction it catalyzes. Once one reaction has been catalyzed, the enzyme is able to participate in other reactions.

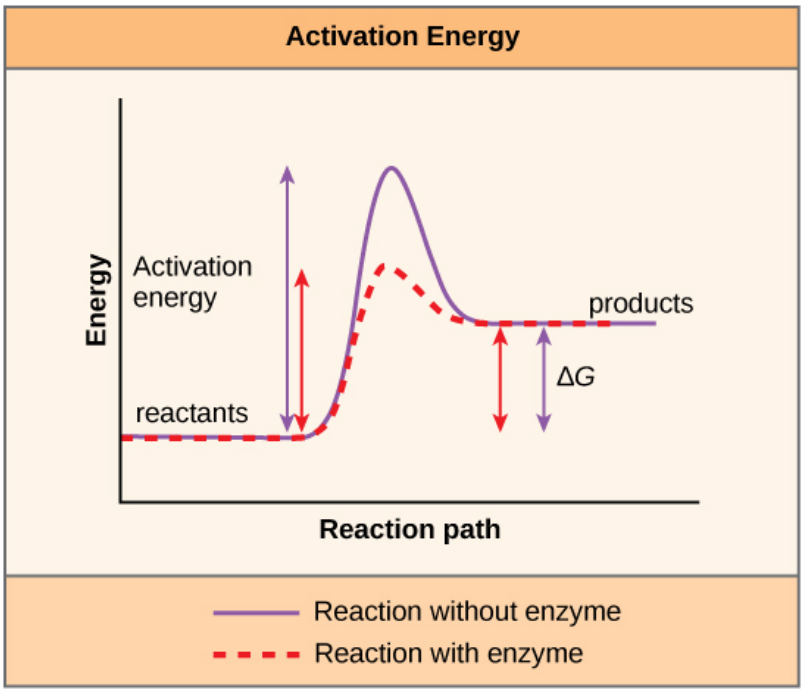


Figure 4.7 Enzymes lower the activation energy of the reaction but do not change the free energy of the reaction.

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

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

For many years, scientists thought that enzyme-substrate binding took place in a simple "lock and key" fashion. This model asserted that the enzyme and substrate fit together perfectly in one instantaneous step. However, current research supports a model called induced fit (Figure 4.8). The induced-fit model expands on the lock-and-key model by describing a more dynamic binding between enzyme and substrate. As the enzyme and substrate come together, their interaction causes a mild shift in the enzyme's structure that forms an ideal binding arrangement between enzyme and substrate.

Concept in Action

View an animation of induced fit.

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

The enzyme-substrate complex can also lower activation energy by compromising the bond structure so that it is easier to break. Finally, enzymes can also lower activation energies by taking part in the chemical reaction itself. In these cases, it is important to remember that the enzyme will always return to its original state by the completion of the reaction. One of the hallmark properties of enzymes is that they remain ultimately unchanged by the reactions they catalyze. After an enzyme has catalyzed a reaction, it releases its product(s) and can catalyze a new reaction.

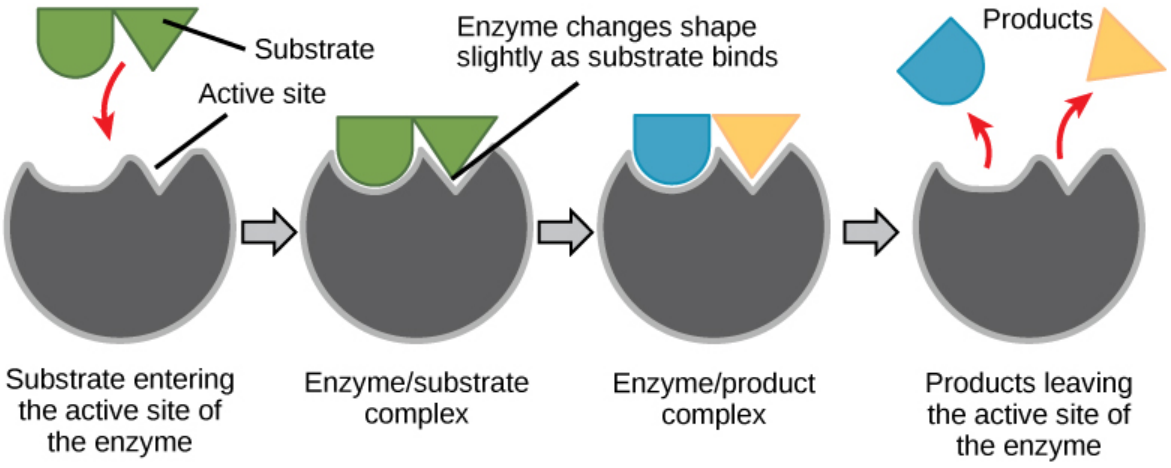


Figure 4.8 The induced-fit model is an adjustment to the lock-and-key model and explains how enzymes and substrates undergo dynamic modifications during the transition state to increase the affinity of the substrate for the active site.

It would seem ideal to have a scenario in which all of an organism's enzymes existed in abundant supply and functioned optimally under all cellular conditions, in all cells, at all times. However, a

variety of mechanisms ensures that this does not happen. Cellular needs and conditions constantly vary from cell to cell, and change within individual cells over time. The required enzymes of stomach cells differ from those of fat storage cells, skin cells, blood cells, and nerve cells. Furthermore, a digestive organ cell works much harder to process and break down nutrients during the time that closely follows a meal compared with many hours after a meal. As these cellular demands and conditions vary, so must the amounts and functionality of different enzymes.

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

Enzymes can also be regulated in ways that either promote or reduce enzyme activity. There are many kinds of molecules that inhibit or promote enzyme function, and various mechanisms by which they do so. In some cases of enzyme **inhibition**, an inhibitor molecule is similar enough to a substrate that it can bind to the active site and simply block the substrate from binding. When this happens, the enzyme is inhibited through **competitive inhibition**, because an inhibitor molecule competes with the substrate for binding to the active site.

On the other hand, in **noncompetitive inhibition**, an inhibitor molecule binds to the enzyme in a location other than the active site, called an **allosteric site**, but still manages to block substrate binding to the active site. Some inhibitor molecules bind to enzymes in a location where their binding induces a conformational change that reduces the affinity of the enzyme for its substrate. This type of inhibition is called allosteric inhibition (Figure 4.9). Most allosterically regulated enzymes are made up of more than one polypeptide, meaning that they have more than one protein subunit. When an allosteric inhibitor binds to a region on an enzyme, all active sites on the protein subunits are changed slightly such that they bind their substrates with less efficiency. There are allosteric activators as well as inhibitors. Allosteric activators bind to locations on an enzyme away from the active site, inducing a conformational change that increases the affinity of the enzyme’s active site(s) for its substrate(s) (Figure 4.9).

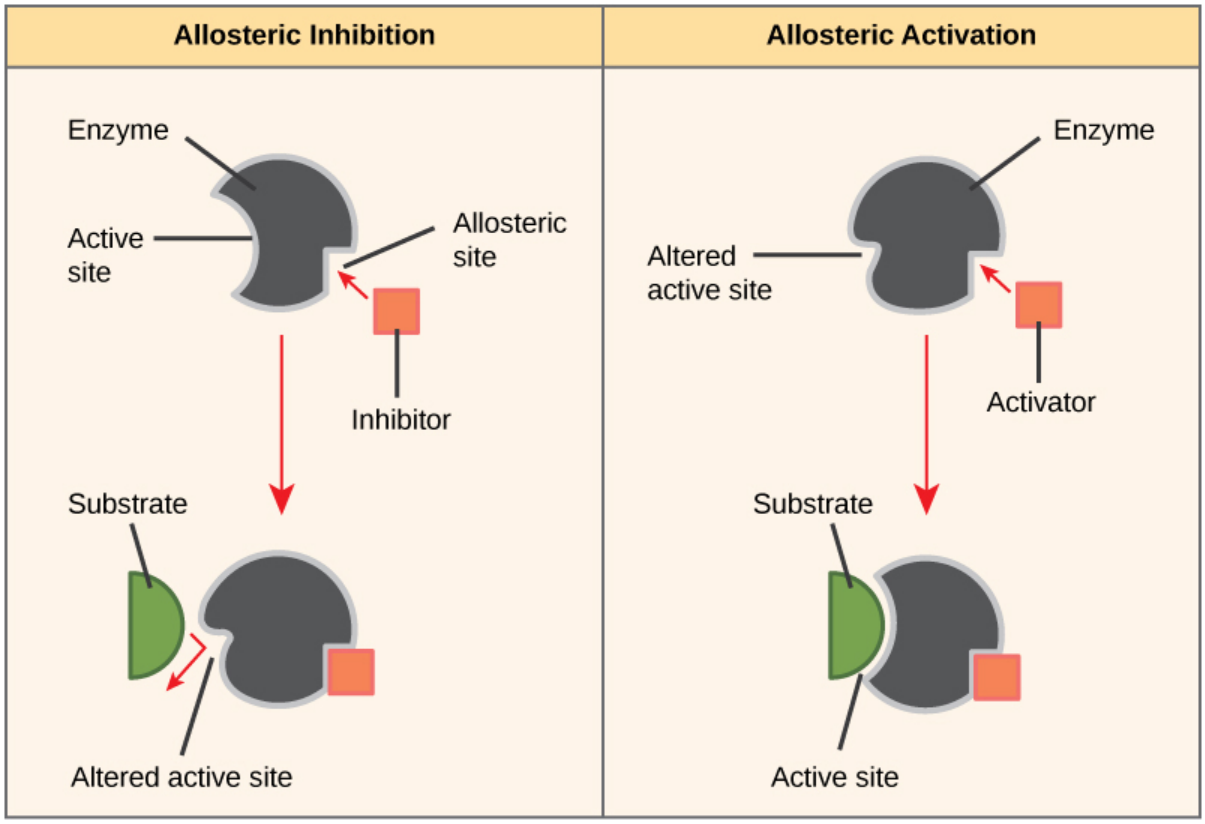


Figure 4.9 Allosteric inhibition works by indirectly inducing a conformational change to the active site such that the substrate no longer fits. In contrast, in allosteric activation, the activator molecule modifies the shape of the active site to allow a better fit of the substrate.

Through the Indigenous Lens

Plants cannot run or hide from their predators and have evolved many strategies to deter those who would eat them. Think of thorns, irritants and secondary metabolites: these are compounds that do not directly help the plant grow, but are made specifically to keep predators away. Secondary metabolites are the most common way plants deter predators. Some examples of secondary metabolites are atropine, nicotine, THC and caffeine. Humans have found these secondary metabolite compounds a rich source of materials for medicines. It is estimated that 90% of the drugs in the modern pharmacy have their “roots” in these secondary metabolites.

First peoples herbal treatments revealed these secondary metabolites to the world. For example, Indigenous peoples have long used the bark of willow shrubs and alder trees for a tea, tonic or poultice to reduce inflammation. You will learn more about the inflammation response by the immune system in chapter 11.

Figure 4.10 Pacific willow bark contains the compound salicin.

Both willow and alder bark contain the compound salicin. Most of us have this compound in our medicine cupboard in the form of salicylic acid or aspirin. Aspirin has been proved to reduce pain and inflammation, and once in our cells salicin converts to salicylic acid.

So how does it work? Salicin or aspirin acts as an enzyme inhibitor. In the inflammatory response two enzymes, COX1 and COX2 are key to this process. Salicin or aspirin specifically modifies an amino acid (serine) in the active site of these two related enzymes. This modification of the active sites does not allow the normal substrate to bind and so the inflammatory process is disrupted. As you have read in this chapter, this makes it competitive enzyme inhibitor.

Pharmaceutical Drug Developer

Figure 4.11 Have you ever wondered how pharmaceutical drugs are developed? (credit: Deborah Austin)

Enzymes are key components of metabolic pathways. Understanding how enzymes work and how they can be regulated are key principles behind the development of many of the pharmaceutical drugs on the market today. Biologists working in this field collaborate with other scientists to design drugs (Figure 4.11).

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

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

Many enzymes do not work optimally, or even at all, unless bound to other specific non-protein helper molecules. They may bond either temporarily through ionic or hydrogen bonds, or permanently through stronger covalent bonds. Binding to these molecules promotes optimal shape and function of their respective enzymes. Two examples of these types of helper molecules are cofactors and

coenzymes. Cofactors are inorganic ions such as ions of iron and magnesium. Coenzymes are organic helper molecules, those with a basic atomic structure made up of carbon and hydrogen. Like enzymes, these molecules participate in reactions without being changed themselves and are ultimately recycled and reused. Vitamins are the source of coenzymes. Some vitamins are the precursors of coenzymes and others act directly as coenzymes. Vitamin C is a direct coenzyme for multiple enzymes that take part in building the important connective tissue, collagen. Therefore, enzyme function is, in part, regulated by the abundance of various cofactors and coenzymes, which may be supplied by an organism's diet or, in some cases, produced by the organism.

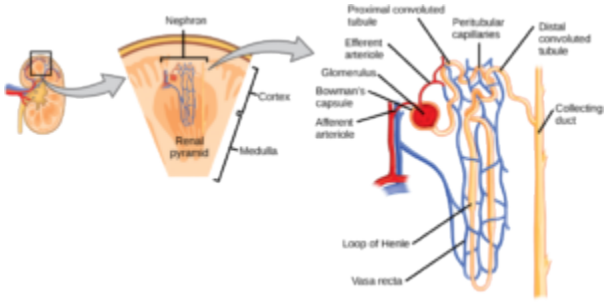


Figure 4.12 Vitamins are important coenzymes or precursors of coenzymes, and are required for enzymes to function properly. Multivitamin capsules usually contain mixtures of all the vitamins at different percentages.

Feedback Inhibition in Metabolic Pathways

Molecules can regulate enzyme function in many ways. The major question remains, however: What are these molecules and where do they come from? Some are cofactors and coenzymes, as you have learned. What other molecules in the cell provide enzymatic regulation such as allosteric modulation, and competitive and non-competitive inhibition? Perhaps the most relevant sources of regulatory molecules, with respect to enzymatic cellular metabolism, are the products of the cellular metabolic reactions themselves. In a most efficient and elegant way, cells have evolved to use the products of their own reactions for feedback inhibition of enzyme activity. Feedback inhibition involves the use of a reaction product to regulate its own further production (Figure 4.12). The cell responds to an abundance of the products by slowing down production during anabolic or catabolic reactions. Such reaction products may inhibit the enzymes that catalyzed their production through the mechanisms described above.

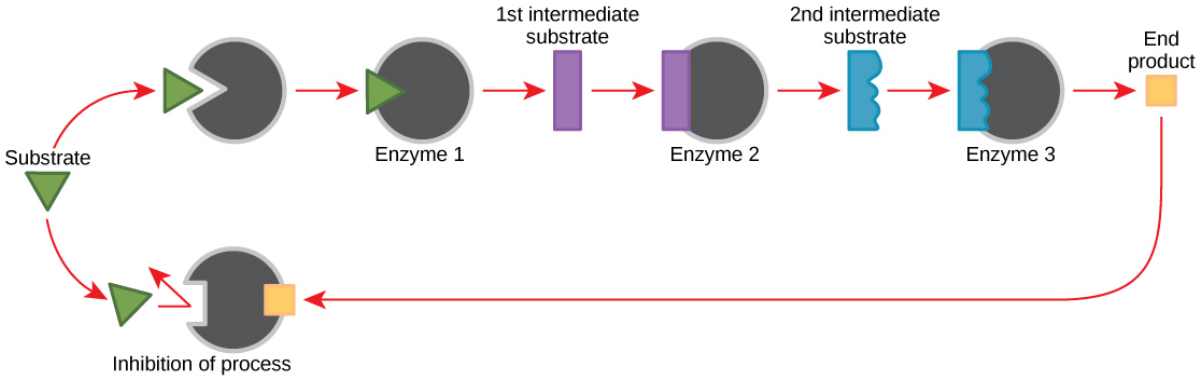


Figure 4.13 Metabolic pathways are a series of reactions catalyzed by multiple enzymes. Feedback inhibition, where the end product of the pathway inhibits an upstream process, is an important regulatory mechanism in cells.

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

Section Summary

Cells perform the functions of life through various chemical reactions. A cell's metabolism refers to the combination of chemical reactions that take place within it. Catabolic reactions break down complex chemicals into simpler ones and are associated with energy release. Anabolic processes build complex molecules out of simpler ones and require energy.

In studying energy, the term system refers to the matter and environment involved in energy transfers. Entropy is a measure of the disorder of a system. The physical laws that describe the transfer of energy are the laws of thermodynamics. The first law states that the total amount of energy in the universe is constant. The second law of thermodynamics states that every energy transfer involves some loss of energy in an unusable form, such as heat energy. Energy comes in different forms: kinetic, potential, and free. The change in free energy of a reaction can be negative (releases energy, exergonic) or positive (consumes energy, endergonic). All reactions require an initial input of energy to proceed, called the activation energy.

Enzymes are chemical catalysts that speed up chemical reactions by lowering their activation energy. Enzymes have an active site with a unique chemical environment that fits particular chemical reactants for that enzyme, called substrates. Enzymes and substrates are thought to bind according to an induced-fit model. Enzyme action is regulated to conserve resources and respond optimally to the environment.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4435#h5p-29

Glossary

activation energy: the amount of initial energy necessary for reactions to occur

active site: a specific region on the enzyme where the substrate binds

allosteric inhibition: the mechanism for inhibiting enzyme action in which a regulatory molecule binds to a second site (not the active site) and initiates a conformation change in the active site, preventing binding with the substrate

anabolic: describes the pathway that requires a net energy input to synthesize complex molecules from simpler ones

bioenergetics: the concept of energy flow through living systems

catabolic: describes the pathway in which complex molecules are broken down into simpler ones, yielding energy as an additional product of the reaction

competitive inhibition: a general mechanism of enzyme activity regulation in which a molecule other than the enzyme's substrate is able to bind the active site and prevent the substrate itself from binding, thus inhibiting the overall rate of reaction for the enzyme

endergonic: describes a chemical reaction that results in products that store more chemical potential energy than the reactants

enzyme: a molecule that catalyzes a biochemical reaction

exergonic: describes a chemical reaction that results in products with less chemical potential energy than the reactants, plus the release of free energy

feedback inhibition: a mechanism of enzyme activity regulation in which the product of a reaction or the final product of a series of sequential reactions inhibits an enzyme for an earlier step in the reaction series

heat energy: the energy transferred from one system to another that is not work

kinetic energy: the type of energy associated with objects in motion

metabolism: all the chemical reactions that take place inside cells, including those that use energy and those that release energy

noncompetitive inhibition: a general mechanism of enzyme activity regulation in which a regulatory molecule binds to a site other than the active site and prevents the active site from binding the substrate; thus, the inhibitor molecule does not compete with the substrate for the active site; allosteric inhibition is a form of noncompetitive inhibition

potential energy: the type of energy that refers to the potential to do work

substrate: a molecule on which the enzyme acts

thermodynamics: the science of the relationships between heat, energy, and work

4.2 Glycolysis

Learning Objectives

By the end of this section, you will be able to:

- Explain how ATP is used by the cell as an energy source
- Describe the overall result in terms of molecules produced of the breakdown of glucose by glycolysis

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). 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 chapter 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.

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 4.13) 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:

RH (Reducing Agent) + NAD^+ (Oxidizing Agent) —-> NADH (Reduced) + R (Oxidized)

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 is 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.

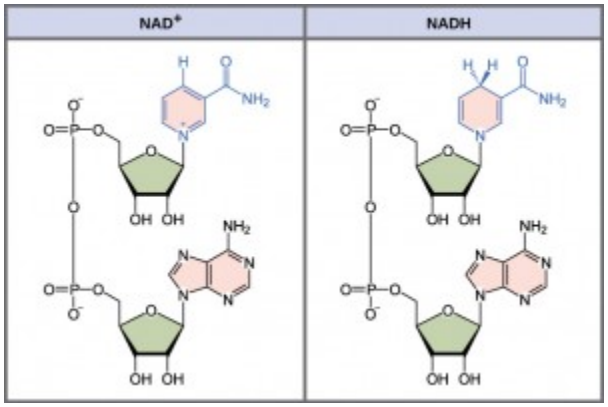


Figure 4.13 The oxidized form of the electron carrier (NAD+) is shown on the left and the reduced form (NADH) is shown on the right. The nitrogenous base in NADH has one more hydrogen ion and two more electrons than in NAD+.

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 the 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 cell uses the energy to do work, 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 cell 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 a single phosphate group (Figure 4.14). 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 to this core molecule results in the formation of adenosine diphosphate (ADP); the addition of a third phosphate group forms adenosine triphosphate (ATP).

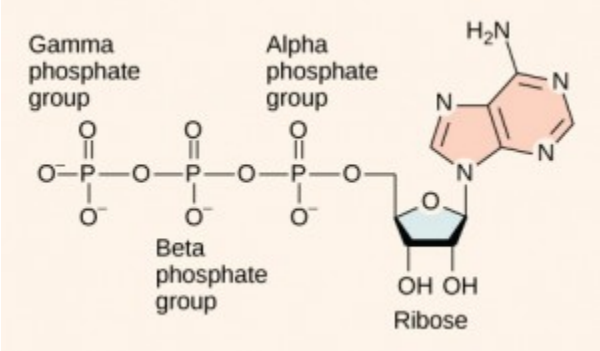


Figure 4.14 ATP (adenosine triphosphate) has three phosphate groups that can be removed by hydrolysis to form ADP (adenosine diphosphate) or AMP (adenosine monophosphate).The negative charges on the phosphate group naturally repel each other, requiring energy to bond them together and releasing energy 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 dephosphorylation, releases energy.

Even exergonic, energy-releasing reactions require a small amount of activation energy to proceed. However, consider endergonic reactions, which require much more energy input because their products have more free energy than their reactants. Within the cell, where does energy to power such reactions come from? The answer lies with an energy-supplying molecule called adenosine triphosphate, or ATP. ATP is a small, relatively simple molecule, but within its bonds contains the potential for a **quick burst of energy that can be harnessed to perform cellular work**. This molecule can be thought of as the primary energy currency of cells in the same way that money is the currency that people exchange for things they need. ATP is used to power the majority of energy-requiring cellular reactions.

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 denature enzymes and other proteins, and thus destroy the cell.

Rather, a cell must be able to store energy safely and release it for use only as needed. Living cells accomplish this using ATP, which can be used to fill any energy need of the cell. How? It functions as a rechargeable battery.

When ATP is broken down, usually by the removal of its terminal phosphate group, energy is released. This energy is used to do work by the cell, usually by the binding of the released phosphate to another molecule, thus activating it. For example, in the mechanical work of muscle contraction, ATP supplies energy to move the contractile muscle proteins.

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 both a ribose molecule and a single phosphate group (Figure 4.15). 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 to this core molecule results in adenosine diphosphate (ADP); the addition of a third phosphate group forms adenosine triphosphate (ATP).

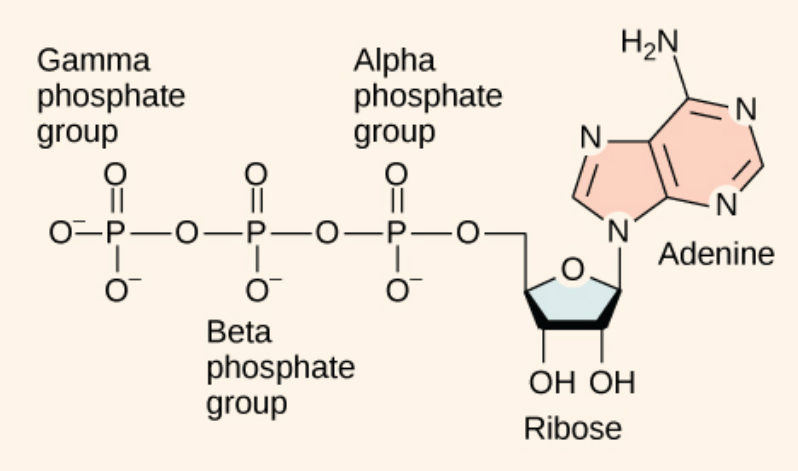


Figure 4.15 The structure of ATP shows the basic components of a two-ring adenine, five-carbon ribose, and three phosphate groups.

The addition of a phosphate group to a molecule requires a high amount of energy and results in a high-energy bond. 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.

Glycolysis

You have read that nearly all of the **energy used by living things comes to them in the bonds of the sugar, glucose**. Glycolysis is the first step in the breakdown of glucose to extract energy for cell metabolism. Many living organisms carry out glycolysis as part of their metabolism. Glycolysis takes place in the cytoplasm of most prokaryotic and all 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. In the first part of the glycolysis pathway, energy is used to make adjustments so that the six-carbon sugar molecule can be split evenly into two three-carbon pyruvate molecules. In the second part of glycolysis, ATP and nicotinamide-adenine dinucleotide (NADH) are produced (Figure 4.16).

If the cell cannot catabolize the pyruvate molecules further, it will harvest only two ATP molecules from one molecule of glucose. For example, mature mammalian red blood cells are only capable of glycolysis, which is their sole source of ATP. If glycolysis is interrupted, these cells would eventually die.

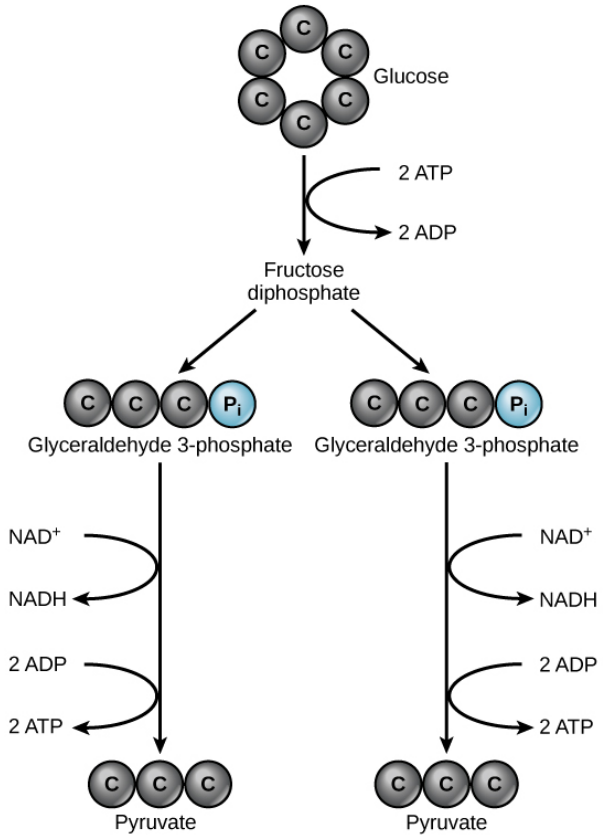


Figure 4.16 In glycolysis, a glucose molecule is converted into two pyruvate molecules.

Section Summary

ATP functions as the energy currency for cells. It allows cells to store energy briefly and transport it within itself to support endergonic chemical reactions. The structure of ATP is that of an RNA nucleotide with three phosphate groups attached. As ATP is used for energy, a phosphate group is detached, and ADP is produced. Energy derived from glucose catabolism is used to recharge ADP into ATP.

Glycolysis is the first pathway used in the breakdown of glucose to extract energy. Because it is used

by nearly all organisms on earth, it must have evolved early in the history of life. Glycolysis consists of two parts: The first part prepares the six-carbon ring of glucose for separation into two three-carbon sugars. Energy from ATP is invested into the molecule during this step to energize the separation. The second half of glycolysis extracts ATP and high-energy electrons from hydrogen atoms and attaches them to NAD^+. Two ATP molecules are invested in the first half and four ATP molecules are formed during the second half. This produces a net gain of two ATP molecules per molecule of glucose for the cell.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4441#h5p-30

Glossary

ATP: (also, adenosine triphosphate) the cell's energy currency

glycolysis: the process of breaking glucose into two three-carbon molecules with the production of ATP and NADH

4.3 Citric Acid Cycle and Oxidative Phosphorylation

Learning Objectives

By the end of this section, you will be able to:

- Describe the location of the citric acid cycle and oxidative phosphorylation in the cell
- Describe the overall outcome of the citric acid cycle and oxidative phosphorylation in terms of the products of each
- Describe the relationships of glycolysis, the citric acid cycle, and oxidative phosphorylation in terms of their inputs and outputs.

The Citric Acid Cycle

In eukaryotic cells, the pyruvate molecules produced at the end of glycolysis are transported into **mitochondria**, which are sites of cellular respiration. If oxygen is available, aerobic respiration will go forward. In mitochondria, pyruvate will be transformed into a two-carbon acetyl group (by removing a molecule of carbon dioxide) that will be picked up by a carrier compound called coenzyme A (CoA), which is made from vitamin B_5. The resulting compound is called acetyl CoA. (Figure 4.17). Acetyl CoA can be used in a variety of ways by the cell, but its major function is to deliver the acetyl group derived from pyruvate to the next pathway in glucose catabolism.

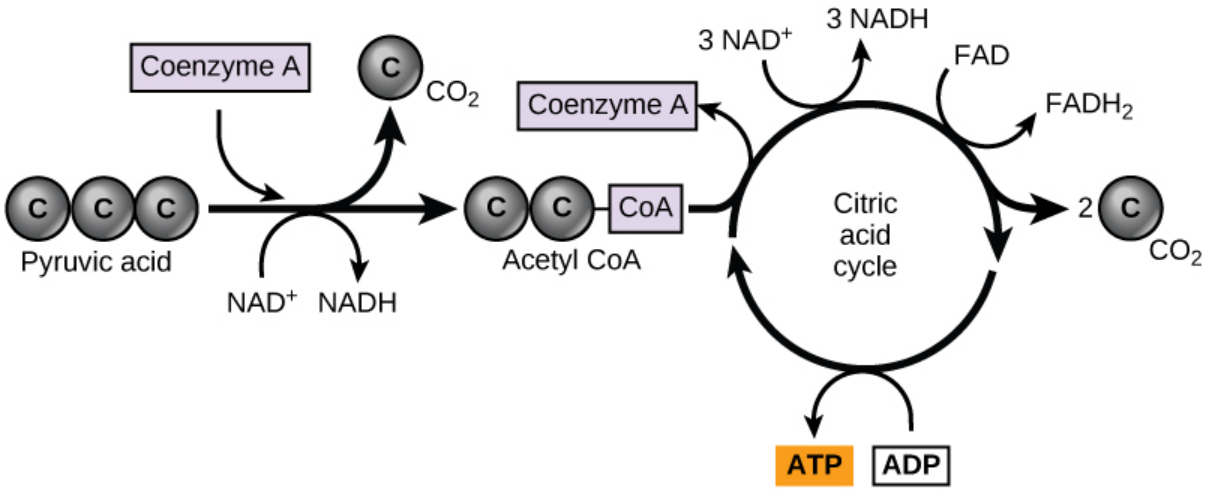


Figure 4.17 Pyruvate is converted into acetyl-CoA before entering the citric acid cycle.

Like the conversion of pyruvate to acetyl CoA, the citric acid cycle in eukaryotic cells takes place in

the **matrix of the mitochondria**. Unlike glycolysis, the citric acid cycle is a closed loop: The last part of the pathway regenerates the compound used in the first step. The eight steps of the cycle are a series of chemical reactions that produces two carbon dioxide molecules, one ATP molecule (or an equivalent), and reduced forms (NADH and $FADH_2$) of NAD^+ and FAD^+, important coenzymes in the cell. Part of this is considered an aerobic pathway (oxygen-requiring) because the NADH and $FADH_2$ produced must transfer their electrons to the next pathway in the system, which will use oxygen. If oxygen is not present, this transfer does not occur.

Two carbon atoms come into the citric acid cycle from each acetyl group. Two carbon dioxide molecules are released on each turn of the cycle; however, these do not contain the same carbon atoms contributed by the acetyl group on that turn of the pathway. The two acetyl-carbon atoms will eventually be released on later turns of the cycle; in this way, all six carbon atoms from the original glucose molecule will be eventually released as carbon dioxide. It takes two turns of the cycle to process the equivalent of one glucose molecule. Each turn of the cycle forms three high-energy NADH molecules and one high-energy $FADH_2$ molecule. These high-energy carriers will connect with the last portion of aerobic respiration to produce ATP molecules. One ATP (or an equivalent) is also made in each cycle. Several of the intermediate compounds in the citric acid cycle can be used in synthesizing non-essential amino acids; therefore, the cycle is both anabolic and catabolic.

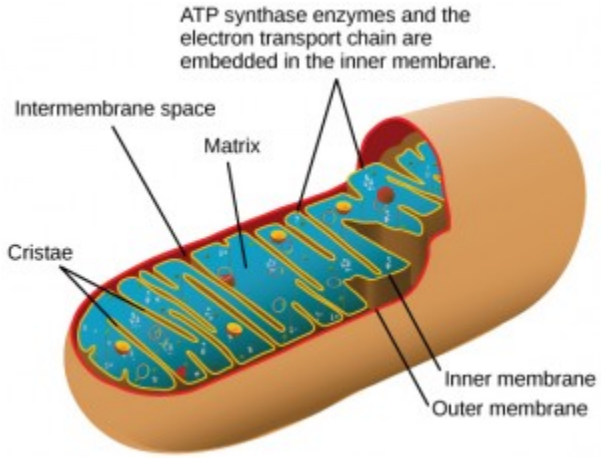


Figure 4.18 In eukaryotes, oxidative phosphorylation takes place in mitochondria. In prokaryotes, this process takes place in the plasma membrane. (Credit: modification of work by Mariana Ruiz Villareal)

Oxidative Phosphorylation

You have just read about two pathways in glucose catabolism—glycolysis and the citric acid 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 derives from a process that begins with passing electrons through a series of chemical reactions to a final electron acceptor, oxygen. These reactions take place in specialized protein complexes located in **the inner membrane of the mitochondria** of eukaryotic organisms and on the inner part of the cell membrane of prokaryotic organisms. The energy of the electrons is harvested and used to generate a **electrochemical gradient**

across the inner mitochondrial membrane. The **potential energy of this gradient is used to generate ATP**. The entirety of this process is called oxidative phosphorylation.

The electron transport chain (Figure 4.19 **a**) is the last component of aerobic respiration and is the only part of metabolism that uses atmospheric oxygen. Oxygen continuously diffuses into plants for this purpose. In animals, oxygen enters the body through the respiratory system. Electron transport is a series of chemical reactions that resembles a bucket brigade in that electrons are passed rapidly from one component to the next, to the endpoint of the chain where oxygen is the final electron acceptor and water is produced. There are four complexes composed of proteins, labeled I through IV in Figure 4.19 **c**, 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 in the plasma membrane of prokaryotes. In each transfer of an electron through the electron transport chain, the electron loses energy, but with some transfers, the energy is stored as potential energy by using it to pump hydrogen ions across the inner mitochondrial membrane into the intermembrane space, creating an electrochemical gradient.

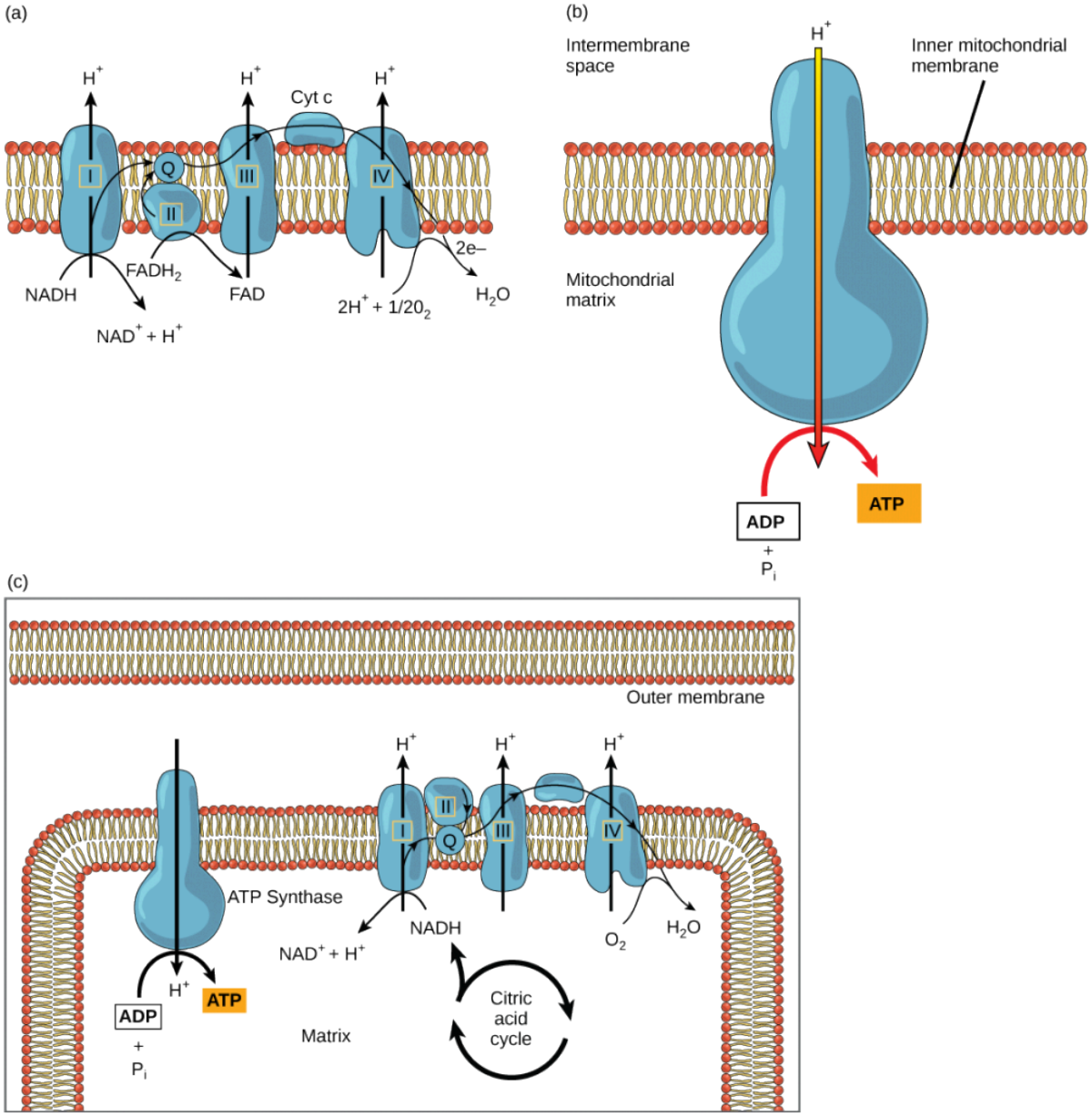


Figure 4.19 (a) The electron transport chain is a set of molecules that supports a series of oxidation-reduction reactions. (b) ATP synthase is a complex, molecular machine that uses an H+ gradient to regenerate ATP from ADP. (c) Chemiosmosis relies on the potential energy provided by the H+ gradient across the membrane.

Cyanide inhibits cytochrome c oxidase, a component of the electron transport chain. If cyanide poisoning occurs, would you expect the pH of the intermembrane space to increase or decrease? What affect would cyanide have on ATP synthesis?

Electrons from NADH and $FADH_2$ are passed to protein complexes in the electron transport chain. As they are passed from one complex to another (there are a total of four), the electrons lose energy, and some of that energy is used to pump hydrogen ions from the mitochondrial matrix into the intermembrane space. In the fourth protein complex, the electrons are accepted by oxygen, the terminal

acceptor. The oxygen with its extra electrons then combines with two hydrogen ions, further enhancing the electrochemical gradient, to form water. If there were no oxygen present in the mitochondrion, the electrons could not be removed from the system, and the entire electron transport chain would back up and stop. The mitochondria would be unable to generate new ATP in this way, and the cell would ultimately die from lack of energy. This is the reason we must breathe to draw in new oxygen.

In the electron transport chain, the free energy from the series of reactions just described is used to pump hydrogen ions across the membrane. The uneven distribution of H^+ ions across the membrane establishes an electrochemical gradient, owing to the H^+ ions' positive charge and their higher concentration on one side of the membrane.

Hydrogen ions diffuse through the inner membrane through an integral membrane protein called ATP synthase (Figure 4.19 **b**). This complex protein acts as a tiny generator, turned by the force of the hydrogen ions diffusing through it, down their electrochemical gradient from the intermembrane space, where there are many mutually repelling hydrogen ions to the matrix, where there are few. The turning of the parts of this molecular machine regenerate ATP from ADP. This flow of hydrogen ions across the membrane through ATP synthase is called chemiosmosis.

Chemiosmosis (Figure 4.19 **c**) is used to generate 90 percent of the ATP made during aerobic glucose catabolism. The result of the reactions is the production of ATP from the energy of the electrons removed from hydrogen atoms. These atoms were originally part of a glucose molecule. At the end of the electron transport system, the electrons are used to reduce an oxygen molecule to oxygen ions. The extra electrons on the oxygen ions attract hydrogen ions (protons) from the surrounding medium, and water is formed. The electron transport chain and the production of ATP through chemiosmosis are collectively called oxidative phosphorylation.

ATP Yield

The number of ATP molecules generated from the catabolism of glucose varies. 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 mitochondrial membrane. The NADH generated from glycolysis cannot easily enter mitochondria. Thus, electrons are picked up on the inside of the mitochondria by either NAD^+ or FAD^+. Fewer ATP molecules are generated when FAD^+ acts as a carrier. NAD^+ is used as the electron transporter in the liver and FAD^+ in the brain, so ATP yield depends on the tissue being considered.

Another factor that affects the yield of ATP molecules generated from glucose is that intermediate compounds 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 somewhat messier than the ideal situations described thus far. For example, sugars other than glucose are fed into the glycolytic pathway for energy extraction. Other molecules that would otherwise be used to harvest energy in glycolysis or the citric acid cycle may be removed to form nucleic acids, amino acids, lipids, or other compounds. Overall, in living systems, these pathways of glucose catabolism extract about 34 percent of the energy contained in glucose.

Mitochondrial Disease Physician

What happens when the critical reactions of cellular respiration do not proceed correctly? Mitochondrial diseases are genetic disorders of metabolism. Mitochondrial disorders can arise from mutations in nuclear or mitochondrial DNA, and they result in the production of less energy than is normal in body cells. Symptoms of mitochondrial diseases can include muscle weakness, lack of coordination, stroke-like episodes, and loss of vision and hearing. Most affected people are diagnosed in childhood, although there are some adult-onset diseases. Identifying and treating mitochondrial disorders is a specialized medical field. The educational preparation for this profession requires a college education, followed by medical school with a specialization in medical genetics. Medical geneticists can be board certified by the American Board of Medical Genetics and go on to become associated with professional organizations devoted to the study of mitochondrial disease, such as the Mitochondrial Medicine Society and the Society for Inherited Metabolic Disease.

Section Summary

The citric acid cycle is a series of chemical reactions that removes high-energy electrons and uses them in the electron transport chain to generate ATP. One molecule of ATP (or an equivalent) is produced per each turn of the cycle.

The electron transport chain is the portion of aerobic respiration that uses free oxygen as the final electron acceptor for electrons removed from the intermediate compounds in glucose catabolism. The electrons are passed through a series of chemical reactions, with a small amount of free energy used at three points to transport hydrogen ions across the membrane. This contributes to the gradient used in chemiosmosis. As the electrons are passed from NADH or $FADH_2$ down the electron transport chain, they lose energy. The products of the electron transport chain are water and ATP. A number of intermediate compounds can be diverted into the anabolism of other biochemical molecules, such as nucleic acids, non-essential amino acids, sugars, and lipids. These same molecules, except nucleic acids, can serve as energy sources for the glucose pathway.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4446#h5p-31

Glossary

acetyl CoA: the combination of an acetyl group derived from pyruvic acid and coenzyme A which is made from pantothenic acid (a B-group vitamin)

ATP synthase: a membrane-embedded protein complex that regenerates ATP from ADP with energy from protons diffusing through it

chemiosmosis: the movement of hydrogen ions down their electrochemical gradient across a membrane through ATP synthase to generate ATP

citric acid cycle: a series of enzyme-catalyzed chemical reactions of central importance in all living cells that harvests the energy in carbon-carbon bonds of sugar molecules to generate ATP; the citric acid cycle is an aerobic metabolic pathway because it requires oxygen in later reactions to proceed

electron transport chain: a series of four large, multi-protein complexes embedded in the inner mitochondrial membrane that accepts electrons from donor compounds and harvests energy from a series of chemical reactions to generate a hydrogen ion gradient across the membrane

oxidative phosphorylation: the production of ATP by the transfer of electrons down the electron transport chain to create a proton gradient that is used by ATP synthase to add phosphate groups to ADP molecules

4.4 Fermentation

Learning Objectives

By the end of this section, you will be able to:

- Discuss the fundamental difference between anaerobic cellular respiration and fermentation
- Describe the type of fermentation that readily occurs in animal cells and the conditions that initiate that fermentation

In aerobic respiration, the final electron acceptor is an oxygen molecule, O_2. If aerobic respiration occurs, then ATP will be produced using the energy of the high-energy electrons carried by NADH or $FADH_2$ to the electron transport chain. If aerobic respiration does not occur, NADH must be reoxidized to NAD^+ for reuse as an electron carrier for glycolysis 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. In contrast, some living systems use an inorganic molecule as a final electron acceptor; both methods are a type of anaerobic cellular respiration. Anaerobic respiration enables organisms to convert energy for their use in the absence of oxygen.

Lactic Acid Fermentation

The fermentation method used by animals and some bacteria like those in yogurt is lactic acid fermentation (Figure 4.20). This occurs routinely in mammalian red blood cells and in skeletal muscle that has insufficient oxygen supply to allow aerobic respiration to continue (that is, in muscles used to the point of fatigue). In muscles, lactic acid produced by fermentation must be removed by the blood circulation and brought to the liver for further metabolism. The chemical reaction of lactic acid fermentation is the following:

$$\textbf{Pyruvic acid} + \textbf{NADH} \leftrightarrow \textbf{lactic acid} + \textbf{NAD}^+ \textbf{Pyruvic acid} + \textbf{NADH} \leftrightarrow \textbf{lactic acid} + \textbf{NAD}^+$$

The enzyme that catalyzes this reaction is lactate dehydrogenase. The reaction can proceed in either direction, but the left-to-right reaction is inhibited by acidic conditions. This lactic acid build-up causes muscle stiffness and fatigue. Once the lactic acid has been removed from the muscle and is circulated to the liver, it can be converted back to pyruvic acid and further catabolized for energy.

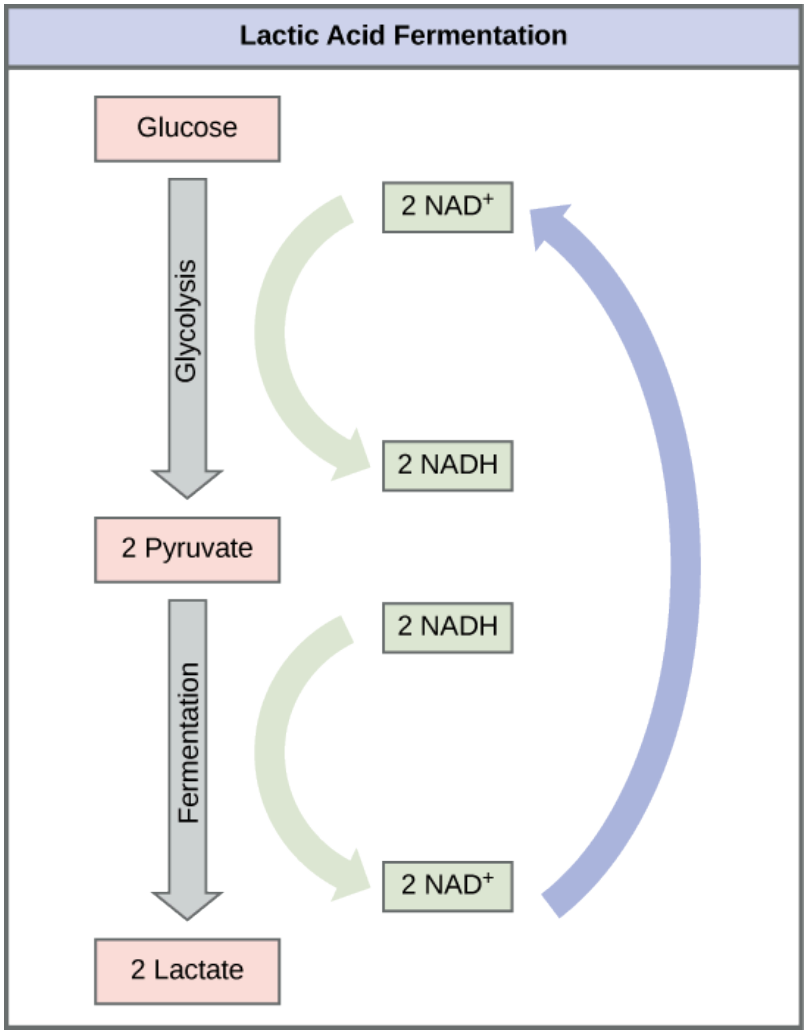


Figure 4.20

Tremetol, a metabolic poison found in white snake root plant, prevents the metabolism of lactate. When cows eat this plant, Tremetol is concentrated in the milk. Humans who consume the milk become ill. Symptoms of this disease, which include vomiting, abdominal pain, and tremors, become worse after exercise. Why do you think this is the case?

<!– The illness is caused by lactic acid build-up. Lactic acid levels rise after exercise, making the symptoms worse. Milk sickness is rare today, but was common in the Midwestern United States in the early 1800s. –>

Alcohol Fermentation

Another familiar fermentation process is alcohol fermentation (Figure 4.21), which produces ethanol, an alcohol. The alcohol fermentation reaction is the following:

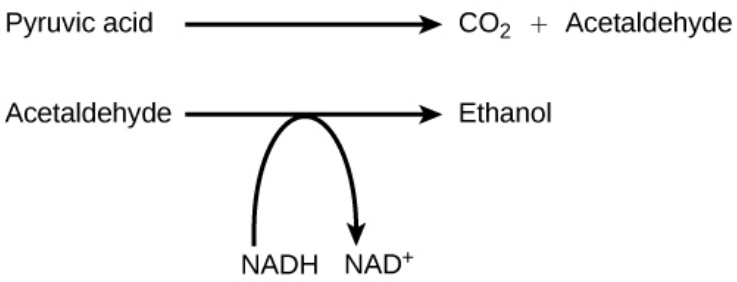


Figure 4.21 The reaction resulting in alcohol fermentation is shown.

In the first reaction, a carboxyl group is removed from pyruvic acid, releasing carbon dioxide as a gas. The loss of carbon dioxide reduces the molecule by one carbon atom, making acetaldehyde. The second reaction removes an electron from NADH, forming NAD^+ and producing ethanol from the acetaldehyde, which accepts the electron. The fermentation of pyruvic acid by yeast produces the ethanol found in alcoholic beverages (Figure 4.22). If the carbon dioxide produced by the reaction is not vented from the fermentation chamber, for example in beer and sparkling wines, it remains dissolved in the medium until the pressure is released. Ethanol above 12 percent is toxic to yeast, so natural levels of alcohol in wine occur at a maximum of 12 percent.

Figure 4.22 Fermentation of grape juice to make wine produces CO2 as a byproduct. Fermentation tanks have valves so that pressure inside the tanks can be released.

Anaerobic Cellular Respiration

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 4.23), reduce sulfate to hydrogen sulfide to regenerate NAD^+ from NADH.

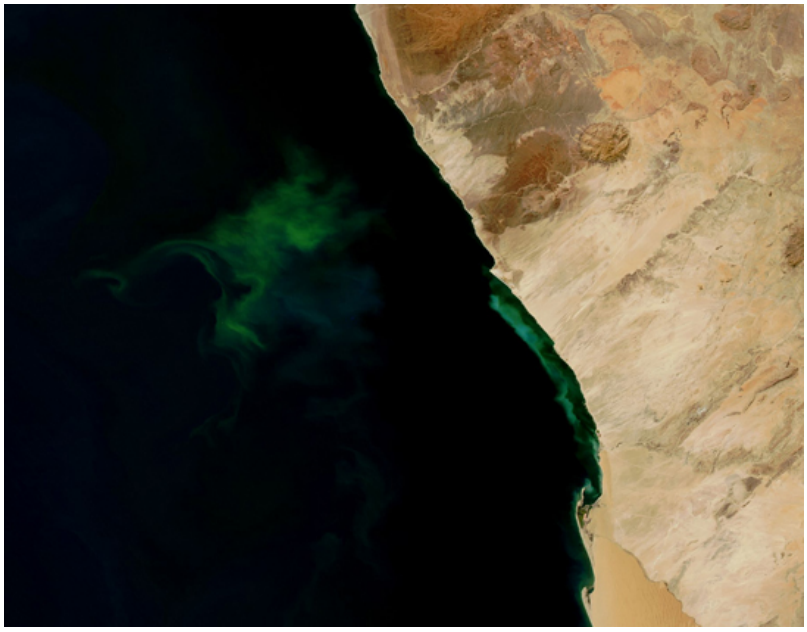


Figure 4.23 The green color seen in these coastal waters is from an eruption of hydrogen sulfide. Anaerobic, sulfate-reducing bacteria release hydrogen sulfide gas as they decompose algae in the water. (credit: NASA image courtesy Jeff Schmaltz, MODIS Land Rapid Response Team at NASA GSFC)

Concept in Action

Visit this site to see anaerobic cellular respiration in action.

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* bacteria, are obligate anaerobes. Obligate anaerobes live and grow in the absence of molecular oxygen. Oxygen is a poison to these microorganisms and kills them upon 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. The various methods of fermentation are used by different 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.

Section Summary

If NADH cannot be metabolized through aerobic respiration, another electron acceptor is used. Most organisms will use some form of fermentation to accomplish the regeneration of NAD^+, ensuring the continuation of glycolysis. The regeneration of NAD^+ in fermentation is not accompanied by ATP production; therefore, the potential for NADH to produce ATP using an electron transport chain is not utilized.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4453#h5p-32

Glossary

anaerobic cellular respiration: the use of an electron acceptor other than oxygen to complete metabolism using electron transport-based chemiosmosis

fermentation: the steps that follow the partial oxidation of glucose via glycolysis to regenerate NAD^+; occurs in the absence of oxygen and uses an organic compound as the final electron acceptor

4.5 Connections to Other Metabolic Pathways

Learning Objectives

By the end of this section, you will be able to:

- Discuss the way in which carbohydrate metabolic pathways, glycolysis, and the citric acid cycle interrelate with protein and lipid metabolic pathways
- Explain why metabolic pathways are not considered closed systems

You have learned about the catabolism of glucose, which provides energy to living cells. But living things consume more than just glucose for food. How does a turkey sandwich, which contains protein, provide energy to your cells? This happens because all of the catabolic pathways for carbohydrates, proteins, and lipids eventually connect into glycolysis and the citric acid cycle pathways (Figure 4.24). Metabolic pathways should be thought of as porous—that is, substances enter from other pathways, and other substances leave for other pathways. These pathways are not closed systems. Many of the products in a particular pathway are reactants in other pathways.

Connections of Other Sugars to Glucose Metabolism

Glycogen, a polymer of glucose, is a short-term energy storage molecule in animals. When there is adequate ATP present, excess glucose is converted into glycogen for storage. **Glycogen is made and stored in the liver and muscle**. Glycogen will be taken out of storage if blood sugar levels drop. The presence of glycogen in muscle cells as a source of glucose allows ATP to be produced for a longer time during exercise.

Sucrose is a disaccharide made from glucose and fructose bonded together. Sucrose is broken down in the small intestine, and the glucose and fructose are absorbed separately. Fructose is one of the three dietary monosaccharides, along with glucose and galactose (which is part of milk sugar, the disaccharide lactose), that 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 broken down by a variety of enzymes in cells. Most of the time, amino acids are recycled into new proteins. If there are excess amino acids, however, or if the body is in a **state of famine**, some amino acids will be shunted into pathways of glucose catabolism. 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 from the nitrogen originating in amino acids, and it leaves the body in urine.

Connections of Lipids to Glucose Metabolism

The lipids that are connected to the glucose pathways are **cholesterol and triglycerides**. Cholesterol is a lipid that contributes to cell membrane flexibility and is a precursor of steroid hormones. The synthesis of cholesterol starts with acetyl CoA and proceeds in only one direction. The process cannot be reversed, and ATP is not produced.

Triglycerides are a form of long-term energy storage in animals. Triglycerides store about twice as much energy as carbohydrates. 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 and proceeds through glycolysis. Fatty acids are broken into two-carbon units that enter the citric acid cycle.

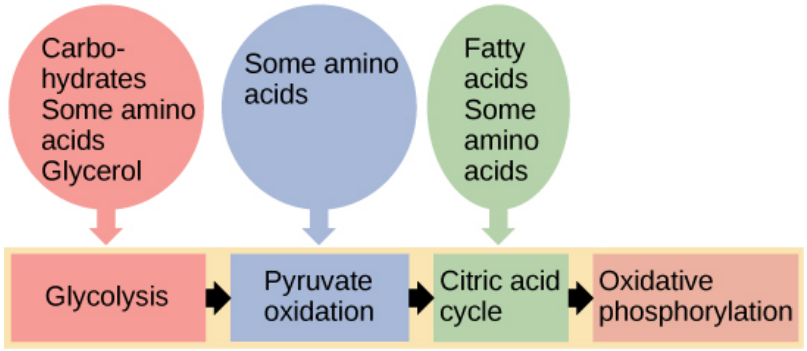


Figure 4.24 Glycogen from the liver and muscles, together with fats, can feed into the catabolic pathways for carbohydrates.

Evolution in Action

Pathways of Photosynthesis and Cellular Metabolism 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. 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 their own cells. 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 that they could use to survive. Additionally, selection would favor those organisms that could extract maximal value from the available nutrients.

An early form of photosynthesis developed that harnessed the sun’s energy using compounds other than water as a source of hydrogen atoms, but this pathway did not produce free oxygen. It is thought that

glycolysis developed prior to this time and could take advantage of simple sugars being produced, but these reactions were not able to fully extract the energy stored in the carbohydrates. A later form of photosynthesis used water as a source of hydrogen ions and generated free oxygen. Over time, the atmosphere became oxygenated. Living things adapted to exploit this new atmosphere and allowed respiration as we know it to evolve. When the full process of photosynthesis as we know it developed and the atmosphere became oxygenated, cells were finally able to use the oxygen expelled by photosynthesis to extract more energy from the sugar molecules using the citric acid cycle.

Section Summary

The breakdown and synthesis of carbohydrates, proteins, and lipids connect with the pathways of glucose catabolism. The carbohydrates that can also feed into glucose catabolism include galactose, fructose, and glycogen. These connect with glycolysis. The amino acids from proteins connect with glucose catabolism through pyruvate, acetyl CoA, and components of the citric acid cycle. Cholesterol synthesis starts with acetyl CoA, and the components of triglycerides are picked up by acetyl CoA and enter the citric acid cycle.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4456#h5p-33

Chapter 5: Introduction to Photosynthesis

Figure 5.1 This sage thrasher's diet, like that of almost all organisms, depends on photosynthesis. (credit: modification of work by Dave Menke, U.S. Fish and Wildlife Service)

No matter how complex or advanced a machine, such as the latest cellular phone, the device cannot function without energy. Living things, similar to machines, have many complex components; they too cannot do anything without energy, which is why humans and all other organisms must "eat" in some form or another. That may be common knowledge, but how many people realize that every bite of every meal ingested depends on the process of photosynthesis?

5.1: Overview of Photosynthesis

Learning Objectives

By the end of this section, you will be able to:

- Summarize the process of photosynthesis
- Explain the relevance of photosynthesis to other living things
- Identify the reactants and products of photosynthesis
- Describe the main structures involved in photosynthesis

All living organisms on earth consist of one or more cells. Each cell 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. Cells then use this energy to perform work, such as cellular respiration.

The energy that is harnessed from photosynthesis enters the ecosystems of our planet continuously and is transferred from one organism to another. Therefore, directly or indirectly, the process of photosynthesis provides most of the energy required by living things on earth.

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.

Concept in Action

Click the following link to learn more about 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 5.2). 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.

Figure 5.2 (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 (Figure 5.3). Using this reasoning, all food eaten by humans also links back to autotrophs that carry out photosynthesis.

Figure 5.3 The energy stored in carbohydrate molecules from photosynthesis passes through the food chain. The predator that eats these deer is getting energy that originated in the photosynthetic vegetation that the deer consumed. (credit: Steve VanRiper, U.S. Fish and Wildlife Service)

Biology in Action

Photosynthesis at the Grocery Store

Figure 5.4 Photosynthesis is the origin of the products that comprise the main elements of the human diet. (credit: Associação Brasileira de Supermercados)

Major grocery stores in the United States are organized into departments, such as dairy, meats, produce,

bread, cereals, and so forth. Each aisle contains hundreds, if not thousands, of different products for customers to buy and consume (Figure 5.4).

Although there is a large variety, each item links back to photosynthesis. Meats and dairy products link to photosynthesis because the animals were fed plant-based foods. The breads, cereals, and pastas come largely from grains, which are the seeds of photosynthetic plants. What about desserts and drinks? All of these products contain sugar—the basic carbohydrate molecule produced directly from photosynthesis. The photosynthesis connection applies to every meal and every food a person consumes.

Main Structures and Summary of Photosynthesis

Photosynthesis requires sunlight, carbon dioxide, and water as starting reactants (Figure 5.5). 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.

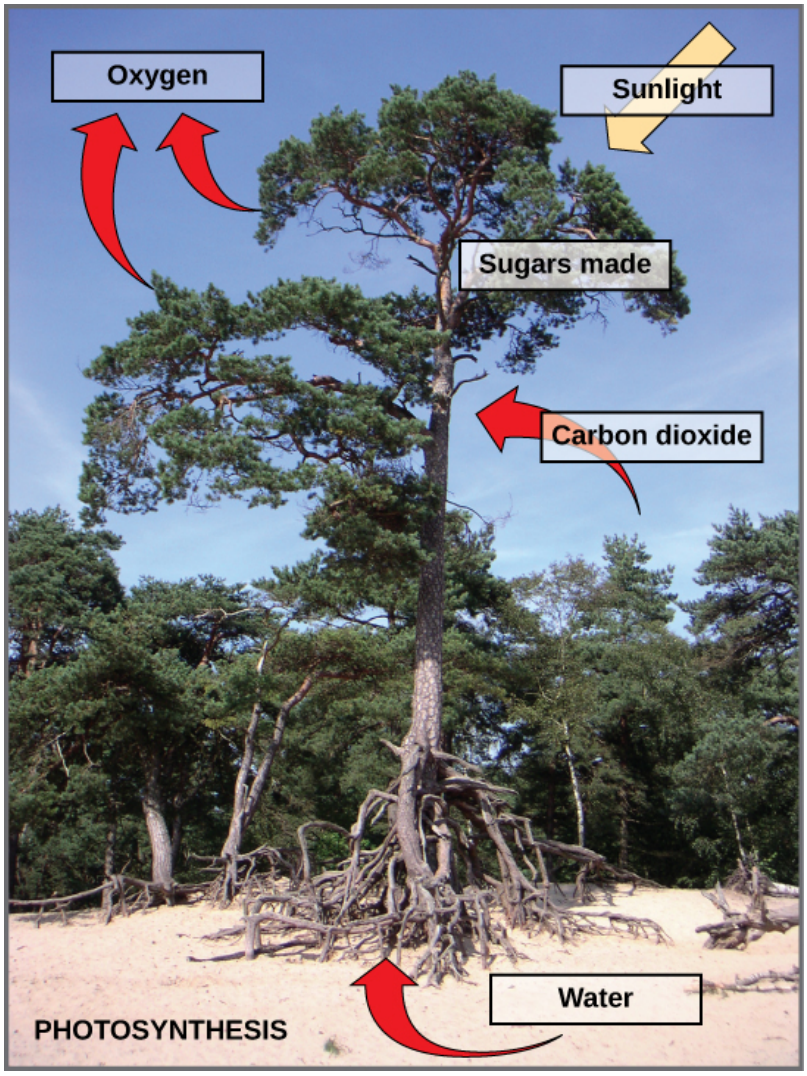


Figure 5.5 Photosynthesis uses solar energy, carbon dioxide, and water to release oxygen to produce energy-storing sugar molecules. Photosynthesis is the origin of the products that comprise the main elements of the human diet. (credit: Associação Brasileira de Supermercados)

The complex reactions of photosynthesis can be summarized by the chemical equation shown in Figure 5.6.

Photosynthesis Equation						
Carbon dioxide	+	Water	SUNLIGHT →	Sugar	+	Oxygen
$6CO_2$		$6H_2O$		$C_6H_{12}O_6$		$6O_2$

Figure 5.6 The process of photosynthesis can be represented by an equation, wherein carbon dioxide and water produce sugar and oxygen using energy from sunlight.

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 5.7). 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 are also involved in photosynthesis, but chlorophyll is by far the most important. As shown in Figure 5.7, a stack of thylakoids is called a granum, and the space surrounding the granum is called stroma (not to be confused with stomata, the openings on the leaves).

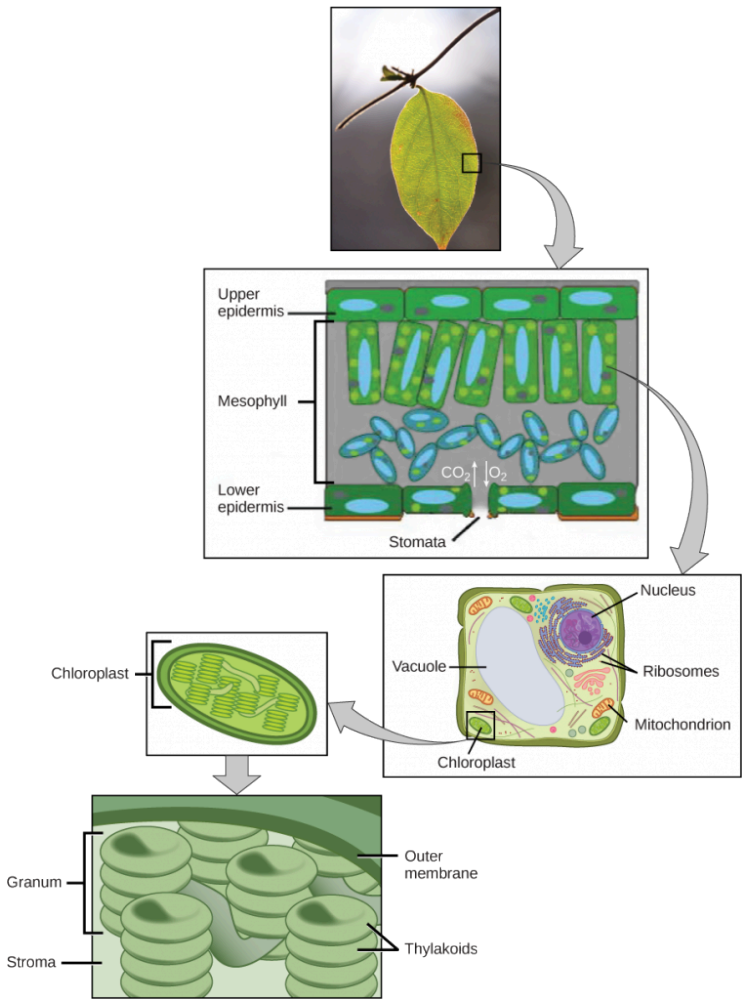


Figure 5.7 Not all cells of a leaf carry out photosynthesis. Cells within the middle layer of a leaf have chloroplasts, which contain the photosynthetic apparatus. (credit "leaf": modification of work by Cory Zanker)

On a hot, dry day, plants close their stomata to conserve water. What impact will this have on photosynthesis?

The Two Parts of Photosynthesis

Photosynthesis takes place in two stages: the light-dependent reactions and the Calvin cycle. 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.

Section Summary

The process of photosynthesis transformed life on earth. By harnessing energy from the sun, photosynthesis allowed living things to access enormous amounts of energy. Because of photosynthesis, living things gained access to sufficient energy, allowing them to evolve new structures and achieve the biodiversity that is evident today.

Only certain organisms, called autotrophs, can perform photosynthesis; they require the presence of chlorophyll, a specialized pigment that can absorb light and convert light energy into chemical energy. Photosynthesis uses carbon dioxide and water to assemble carbohydrate molecules (usually glucose) and releases oxygen into the air. Eukaryotic autotrophs, such as plants and algae, have organelles called chloroplasts in which photosynthesis takes place.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4468#h5p-34

Glossary

autotroph: an organism capable of producing its own food

chlorophyll: the green pigment that captures the light energy that drives the reactions of photosynthesis

chloroplast: the organelle where photosynthesis takes place

granum: a stack of thylakoids located inside a chloroplast

heterotroph: an organism that consumes other organisms for food

light-dependent reaction: the first stage of photosynthesis where visible light is absorbed to form two energy-carrying molecules (ATP and NADPH)

mesophyll: the middle layer of cells in a leaf

photoautotroph: an organism capable of synthesizing its own food molecules (storing energy), using the energy of light

pigment: a molecule that is capable of absorbing light energy

stoma: the opening that regulates gas exchange and water regulation between leaves and the environment; plural: stomata

stroma: the fluid-filled space surrounding the grana inside a chloroplast where the Calvin cycle reactions of photosynthesis take place

thylakoid: a disc-shaped membranous structure inside a chloroplast where the light-dependent reactions of photosynthesis take place using chlorophyll embedded in the membranes

5.2: The Light-Dependent Reactions of Photosynthesis

Learning Objectives

By the end of this section, you will be able to:

- Explain how plants absorb energy from sunlight
- Describe how the wavelength of light affects its energy and color
- Describe how and where photosynthesis takes place within a plant

How can light be used to make food? It is easy to think of light as something that exists and allows living organisms, such as humans, to see, but light is a form of energy. Like all energy, light can travel, change form, and be harnessed to do work. In the case of photosynthesis, light energy is transformed into chemical energy, which autotrophs use to build carbohydrate molecules. However, autotrophs only use a specific component of sunlight (Figure 5.8).

Figure 5.8 Autotrophs can capture light energy from the sun, converting it into chemical energy used to build food molecules. (credit: modification of work by Gerry Atwell, U.S. Fish and Wildlife Service)

Concept in Action

Visit this site and click through the animation to view the process of photosynthesis within a leaf.

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 is referred to as “visible light.” The manner in which solar energy travels can be described and measured as waves. Scientists can determine the amount of energy of a wave by measuring its wavelength, the distance between two consecutive, similar points in a series of waves, such as from crest to crest or trough to trough (Figure 5.9).

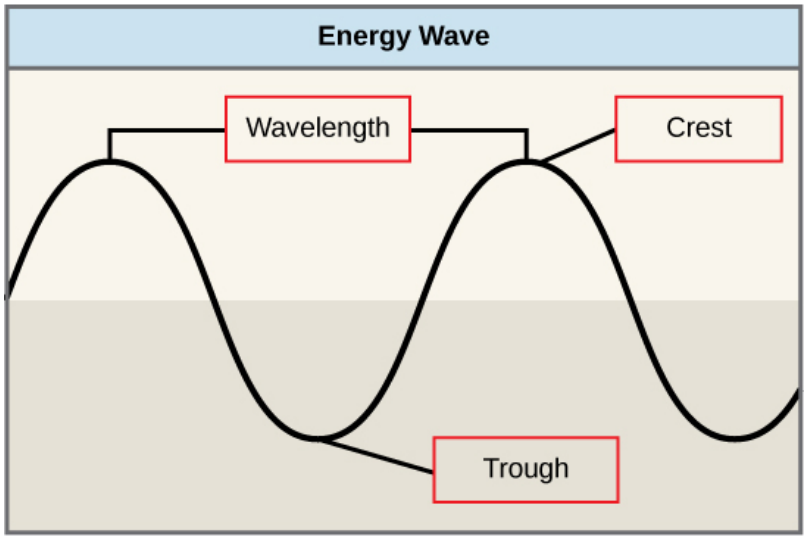


Figure 5.9 The wavelength of a single wave is the distance between two consecutive points along the wave.

Visible light constitutes only one of many types of electromagnetic radiation emitted from the sun. The electromagnetic spectrum is the range of all possible wavelengths of radiation (Figure 5.10). Each wavelength corresponds to a different amount of energy carried.

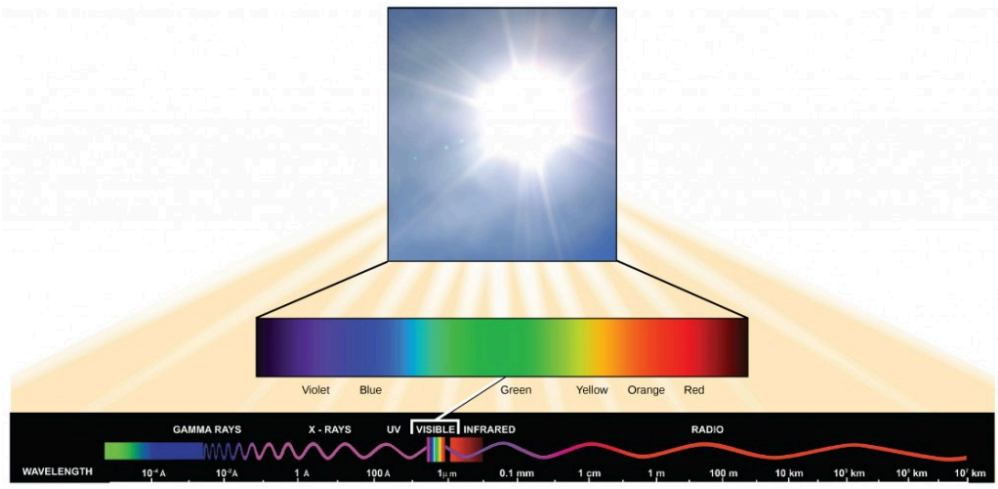


Figure 5.10 The sun emits energy in the form of electromagnetic radiation. This radiation exists in different wavelengths, each of which has its own characteristic energy. Visible light is one type of energy emitted from the sun.

Each type of electromagnetic radiation has a characteristic range of wavelengths. The longer the wavelength (or the more stretched out it appears), 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 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 sun emits a broad range of electromagnetic radiation, including X-rays and ultraviolet (UV) rays. The higher-energy waves are dangerous to living things; for example, X-rays and UV rays can be harmful to humans.

Absorption of Light

Light energy enters the process of photosynthesis when pigments absorb the light. In plants, pigment molecules absorb only visible light for photosynthesis. 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 these colors to the human eye. The visible light portion of the electromagnetic spectrum is perceived by the human eye as a 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.

Understanding Pigments

Different kinds of pigments exist, and each absorbs only certain wavelengths (colors) of visible light. Pigments reflect the color of the wavelengths that they cannot absorb.

All photosynthetic organisms contain a pigment called chlorophyll *a*, which humans see as the common green color associated with plants. Chlorophyll *a* absorbs wavelengths from either end of the

visible spectrum (blue and red), but not from green. Because green is reflected, chlorophyll appears green.

Other pigment types include chlorophyll *b* (which absorbs blue and red-orange light) and the carotenoids. Each type of pigment can be identified by the specific pattern of wavelengths it absorbs from visible light, which is its absorption spectrum.

Many photosynthetic organisms have a mixture of pigments; between them, the organism can absorb energy from a wider range of visible-light wavelengths. Not all photosynthetic organisms have full access to sunlight. Some organisms grow underwater where light intensity decreases with depth, and certain wavelengths are absorbed by the water. 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 block most of the sunlight (Figure 5.11).

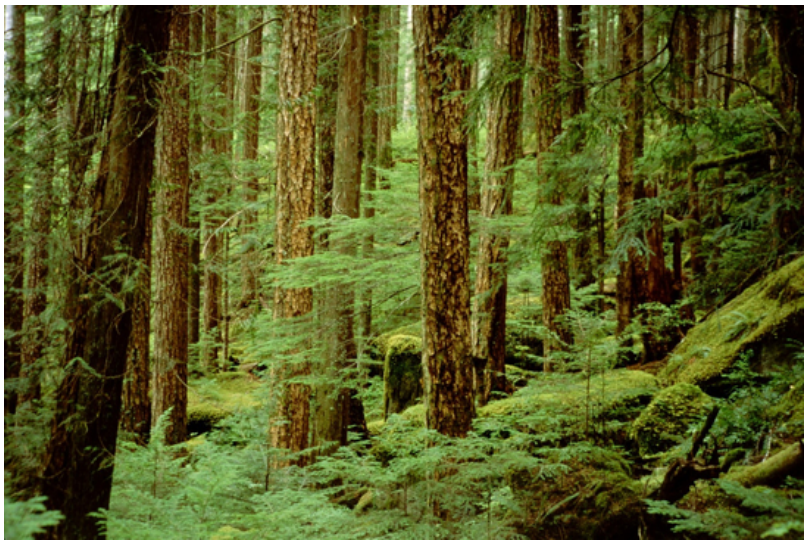


Figure 5.11 Plants that commonly grow in the shade benefit from having a variety of light-absorbing pigments. Each pigment can absorb different wavelengths of light, which allows the plant to absorb any light that passes through the taller trees. (credit: Jason Hollinger)

How Light-Dependent Reactions Work

The overall purpose of the light-dependent reactions is to convert light energy into chemical energy. This chemical energy will be used by the Calvin cycle to fuel the assembly of sugar molecules.

The light-dependent reactions begin in a grouping of pigment molecules and proteins called a photosystem. Photosystems exist in the membranes of thylakoids. A pigment molecule in the photosystem absorbs one photon, a quantity or “packet” of light energy, at a time.

A photon of light energy travels until it reaches a molecule of chlorophyll. The photon causes an electron in the chlorophyll to become “excited.” The energy given to the electron allows it to break free from an atom of the chlorophyll molecule. Chlorophyll is therefore said to “donate” an electron (Figure 5.12).

To replace the electron in the chlorophyll, a molecule of water is split. This splitting releases an electron and results in the formation of oxygen (O_2) and hydrogen ions (H^+) in the thylakoid space. Technically, each breaking of a water molecule releases a pair of electrons, and therefore can replace two donated electrons.

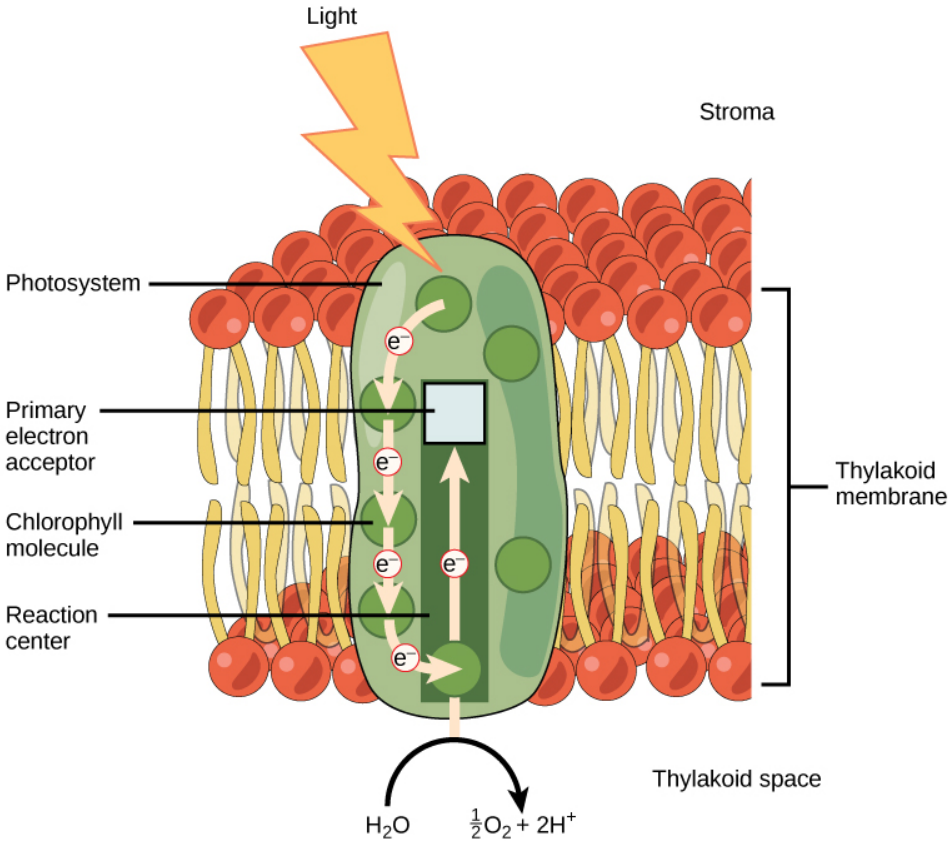


Figure 5.12 Light energy is absorbed by a chlorophyll molecule and is passed along a pathway to other chlorophyll molecules. The energy culminates in a molecule of chlorophyll found in the reaction center. The energy "excites" one of its electrons enough to leave the molecule and be transferred to a nearby primary electron acceptor. A molecule of water splits to release an electron, which is needed to replace the one donated. Oxygen and hydrogen ions are also formed from the splitting of water.

The replacing of the electron enables chlorophyll to respond to another photon. The oxygen molecules produced as byproducts find their way to the surrounding environment. The hydrogen ions play critical roles in the remainder of the light-dependent reactions.

Keep in mind that the purpose of the light-dependent reactions is to convert solar energy into chemical carriers that will be used in the Calvin cycle. In eukaryotes and some prokaryotes, two photosystems exist. The first is called photosystem II, which was named for the order of its discovery rather than for the order of the function.

After the photon hits, photosystem II transfers the free electron to the first in a series of proteins inside the thylakoid membrane called the electron transport chain. As the electron passes along these proteins, energy from the electron fuels membrane pumps that actively move hydrogen ions against their concentration gradient from the stroma into the thylakoid space. This is quite analogous to the process that occurs in the mitochondrion in which an electron transport chain pumps hydrogen ions from the

mitochondrial stroma across the inner membrane and into the intermembrane space, creating an electrochemical gradient. After the energy is used, the electron is accepted by a pigment molecule in the next photosystem, which is called photosystem I (Figure 5.13).

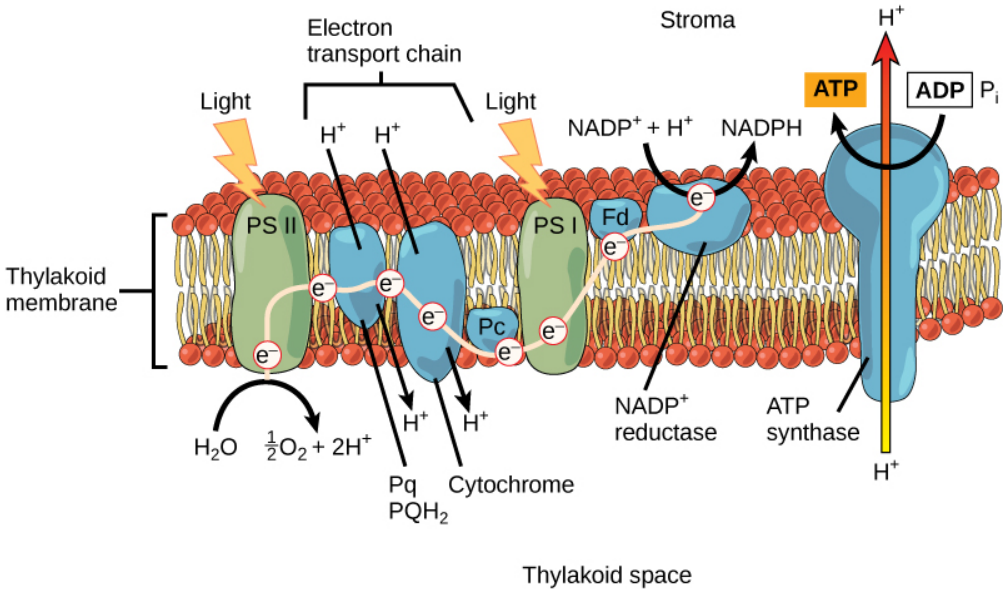


Figure 5.13 From photosystem II, the electron travels along a series of proteins. This electron transport system uses the energy from the electron to pump hydrogen ions into the interior of the thylakoid. A pigment molecule in photosystem I accepts the electron.

Generating an Energy Carrier: ATP

In the light-dependent reactions, energy absorbed by sunlight is stored by two types of energy-carrier molecules: ATP and NADPH. The energy that these molecules carry is stored in a bond that holds a single atom to the molecule. For ATP, it is a phosphate atom, and for NADPH, it is a hydrogen atom. Recall that NADH was a similar molecule that carried energy in the mitochondrion from the citric acid cycle to the electron transport chain. When these molecules release energy into the Calvin cycle, they each lose atoms to become the lower-energy molecules ADP and $NADP^+$.

The buildup of hydrogen ions in the thylakoid space forms an electrochemical gradient because of the difference in the concentration of protons (H^+) and the difference in the charge across the membrane that they create. This potential energy is harvested and stored as chemical energy in ATP through chemiosmosis, the movement of hydrogen ions down their electrochemical gradient through the transmembrane enzyme ATP synthase, just as in the mitochondrion.

The hydrogen ions are allowed to pass through the thylakoid membrane through an embedded protein complex called ATP synthase. This same protein generated ATP from ADP in the mitochondrion. The energy generated by the hydrogen ion stream allows ATP synthase to attach a third phosphate to ADP, which forms a molecule of ATP in a process called photophosphorylation. The flow of hydrogen ions through ATP synthase is called chemiosmosis, because the ions move from an area of high to low concentration through a semi-permeable structure.

Generating Another Energy Carrier: NADPH

The remaining function of the light-dependent reaction is to generate the other energy-carrier molecule, NADPH. As the electron from the electron transport chain arrives at photosystem I, it is re-energized with another photon captured by chlorophyll. The energy from this electron drives the formation of NADPH from $NADP^+$ and a hydrogen ion (H^+). Now that the solar energy is stored in energy carriers, it can be used to make a sugar molecule.

Section Summary

In the first part of photosynthesis, the light-dependent reaction, pigment molecules absorb energy from sunlight. The most common and abundant pigment is chlorophyll *a*. A photon strikes photosystem II to initiate photosynthesis. Energy travels through the electron transport chain, which pumps hydrogen ions into the thylakoid space. This forms an electrochemical gradient. The ions flow through ATP synthase from the thylakoid space into the stroma in a process called chemiosmosis to form molecules of ATP, which are used for the formation of sugar molecules in the second stage of photosynthesis. Photosystem I absorbs a second photon, which results in the formation of an NADPH molecule, another energy carrier for the Calvin cycle reactions.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4477#h5p-35

Glossary

absorption spectrum: the specific pattern of absorption for a substance that absorbs electromagnetic radiation

chlorophyll *a*: the form of chlorophyll that absorbs violet-blue and red light

chlorophyll *b*: the form of chlorophyll that absorbs blue and red-orange light

electromagnetic spectrum: the range of all possible frequencies of radiation

photon: a distinct quantity or “packet” of light energy

photosystem: a group of proteins, chlorophyll, and other pigments that are used in the light-dependent reactions of photosynthesis to absorb light energy and convert it into chemical energy

wavelength: the distance between consecutive points of a wave

5.3: The Calvin Cycle

Learning Objectives

By the end of this section, you will be able to:

- Describe the Calvin cycle
- Define carbon fixation
- Explain how photosynthesis works in the energy cycle of all living organisms

After the energy from the sun is converted and packaged into ATP and NADPH, the cell has the fuel needed to build food in the form of carbohydrate molecules. The carbohydrate molecules made will have a backbone of carbon atoms. Where does the carbon come from? The carbon atoms used to build carbohydrate molecules comes from carbon dioxide, the gas that animals exhale with each breath. The Calvin cycle is the term used for the reactions of photosynthesis that use the energy stored by the light-dependent reactions to form glucose and other carbohydrate molecules.

The Interworkings of the Calvin Cycle

In plants, carbon dioxide (CO_2) enters the chloroplast through the stomata and diffuses into the stroma of the chloroplast—the site of the Calvin cycle reactions where sugar is synthesized. The reactions are named after the scientist who discovered them, and reference the fact that the reactions function as a cycle. Others call it the Calvin-Benson cycle to include the name of another scientist involved in its discovery (Figure 5.14).

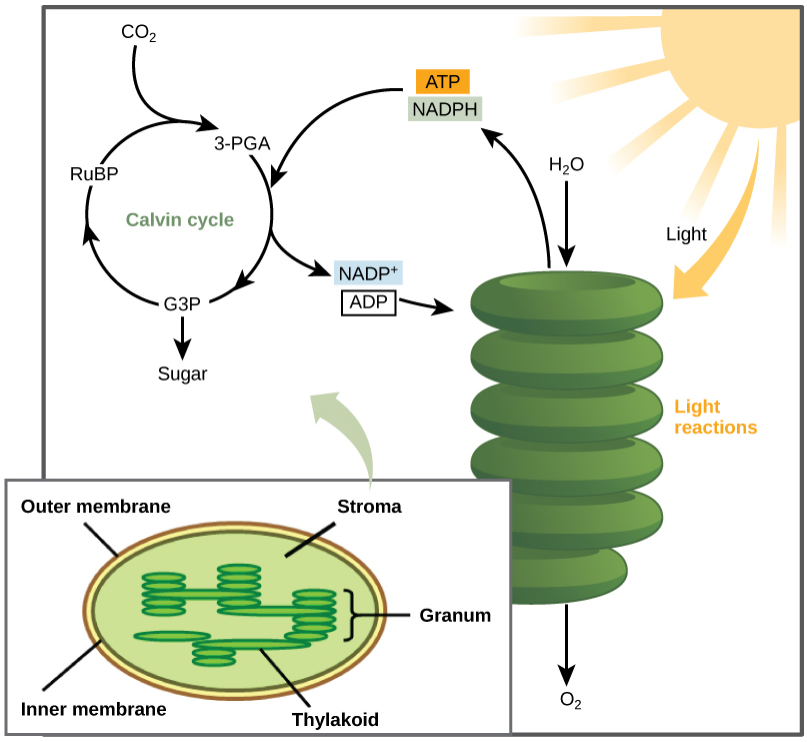


Figure 5.14 Light-dependent reactions harness energy from the sun to produce ATP and NADPH. These energy-carrying molecules travel into the stroma where the Calvin cycle reactions take place.

The Calvin cycle reactions (Figure 5.15) can be organized into three basic stages: fixation, reduction, and regeneration. In the stroma, in addition to CO_2, two other chemicals are present to initiate the Calvin cycle: an enzyme abbreviated RuBisCO, and the molecule ribulose bisphosphate (RuBP). RuBP has five atoms of carbon and a phosphate group on each end.

RuBisCO catalyzes a reaction between CO_2 and RuBP, which forms a six-carbon compound that is immediately converted into two three-carbon compounds. This process is called carbon fixation, because CO_2 is "fixed" from its inorganic form into organic molecules.

ATP and NADPH use their stored energy to convert the three-carbon compound, 3-PGA, into another three-carbon compound called G3P. This type of reaction is called a reduction reaction, because it involves the gain of electrons. A reduction is the gain of an electron by an atom or molecule. The molecules of ADP and NAD^+, resulting from the reduction reaction, return to the light-dependent reactions to be re-energized.

One of the G3P molecules leaves the Calvin cycle to contribute to the formation of the carbohydrate molecule, which is commonly glucose ($C_6H_{12}O_6$). Because the carbohydrate molecule has six carbon atoms, it takes six turns of the Calvin cycle to make one carbohydrate molecule (one for each carbon dioxide molecule fixed). The remaining G3P molecules regenerate RuBP, which enables the system to prepare for the carbon-fixation step. ATP is also used in the regeneration of RuBP.

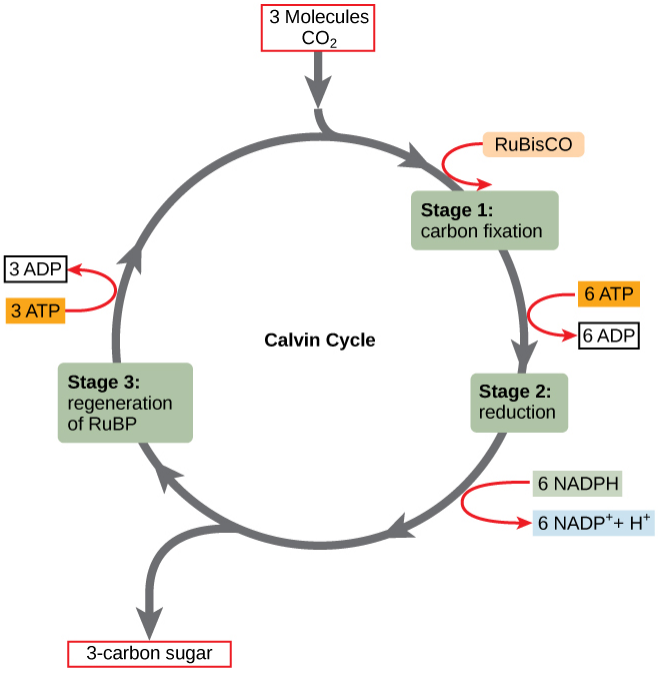


Figure 5.15 The Calvin cycle has three stages. In stage 1, the enzyme RuBisCO incorporates carbon dioxide into an organic molecule. In stage 2, the organic molecule is reduced. In stage 3, RuBP, the molecule that starts the cycle, is regenerated so that the cycle can continue.

In summary, it takes six turns of the Calvin cycle to fix six carbon atoms from CO_2. These six turns require energy input from 12 ATP molecules and 12 NADPH molecules in the reduction step and 6 ATP molecules in the regeneration step.

Concept in Action

The following is a link to an animation of the Calvin cycle. Click Stage 1, Stage 2, and then Stage 3 to see G3P and ATP regenerate to form RuBP.

Photosynthesis

The shared evolutionary history of all photosynthetic organisms is conspicuous, as the basic process has changed little over eras of time. Even between the giant tropical leaves in the rainforest and tiny cyanobacteria, the process and components of photosynthesis that use water as an electron donor

remain largely the same. Photosystems function to absorb light and use electron transport chains to convert energy. The Calvin cycle reactions assemble carbohydrate molecules with this energy.

However, as with all biochemical pathways, a variety of conditions leads to varied adaptations that affect the basic pattern. Photosynthesis in dry-climate plants (Figure 5.16) has evolved with adaptations that conserve water. In the harsh dry heat, every drop of water and precious energy must be used to survive. Two adaptations have evolved in such plants. In one form, a more efficient use of CO_2 allows plants to photosynthesize even when CO_2 is in short supply, as when the stomata are closed on hot days. The other adaptation performs preliminary reactions of the Calvin cycle at night, because opening the stomata at this time conserves water due to cooler temperatures. In addition, this adaptation has allowed plants to carry out low levels of photosynthesis without opening stomata at all, an extreme mechanism to face extremely dry periods.

Figure 5.16 Living in the harsh conditions of the desert has led plants like this cactus to evolve variations in reactions outside the Calvin cycle. These variations increase efficiency and help conserve water and energy. (credit: Piotr Wojtkowski)

Photosynthesis in Prokaryotes

The two parts of photosynthesis—the light-dependent reactions and the Calvin cycle—have been described, as they take place in chloroplasts. However, prokaryotes, such as cyanobacteria, lack membrane-bound organelles. Prokaryotic photosynthetic autotrophic organisms have infoldings of the

plasma membrane for chlorophyll attachment and photosynthesis (Figure 5.17). It is here that organisms like cyanobacteria can carry out photosynthesis.

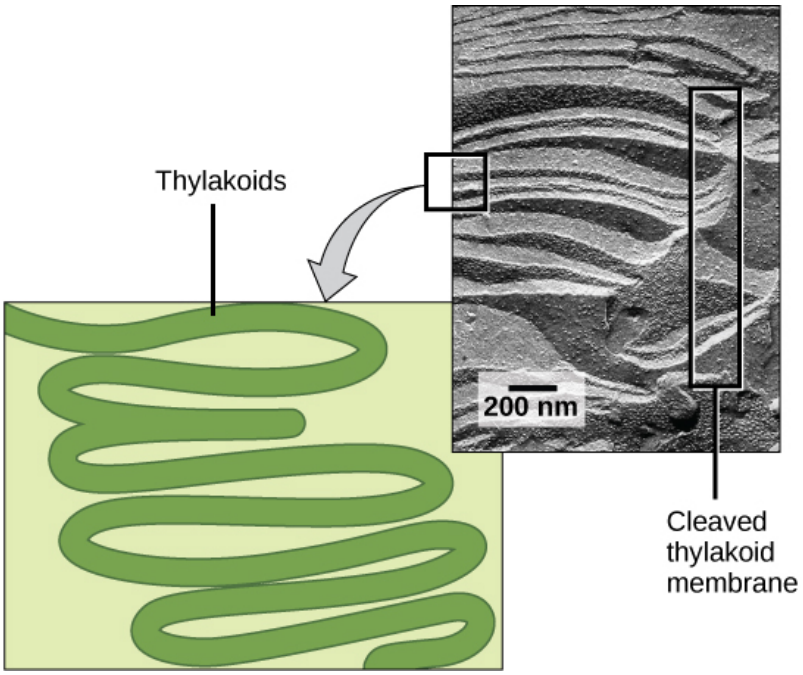


Figure 5.17 A photosynthetic prokaryote has infolded regions of the plasma membrane that function like thylakoids. Although these are not contained in an organelle, such as a chloroplast, all of the necessary components are present to carry out photosynthesis. (credit: scale-bar data from Matt Russell)

The Energy Cycle

Living things access energy by breaking down carbohydrate molecules. However, if plants make carbohydrate molecules, why would they need to break them down? Carbohydrates are storage molecules for energy in all living things. Although energy can be stored in molecules like ATP, carbohydrates are much more stable and efficient reservoirs for chemical energy. Photosynthetic organisms also carry out the reactions of respiration to harvest the energy that they have stored in carbohydrates, for example, plants have mitochondria in addition to chloroplasts.

You may have noticed that the overall reaction for photosynthesis:
$$\mathbf{6CO_2+6H_2O \rightarrow C_{6}{}_{12}O_6+6O_2 6CO_2+6H_2O \rightarrow C_6H_{12}O_6+6O}$$

is the reverse of the overall reaction for cellular respiration:
$$\mathbf{6O_2+C_6H_{12}O_6 \rightarrow 6CO_2+6H_2O_6O_2+C_6H_{12}O_6 \rightarrow 6CO_2+6H_2O}$$

Photosynthesis produces oxygen as a byproduct, and respiration produces carbon dioxide as a byproduct.

In nature, there is no such thing as waste. Every single atom of matter is conserved, recycling

indefinitely. Substances change form or move from one type of molecule to another, but never disappear (Figure 5.18).

CO_2 is no more a form of waste produced by respiration than oxygen is a waste product of photosynthesis. Both are byproducts of reactions that move on to other reactions. Photosynthesis absorbs energy to build carbohydrates in chloroplasts, and aerobic cellular respiration releases energy by using oxygen to break down carbohydrates. Both organelles use electron transport chains to generate the energy necessary to drive other reactions. Photosynthesis and cellular respiration function in a biological cycle, allowing organisms to access life-sustaining energy that originates millions of miles away in a star.

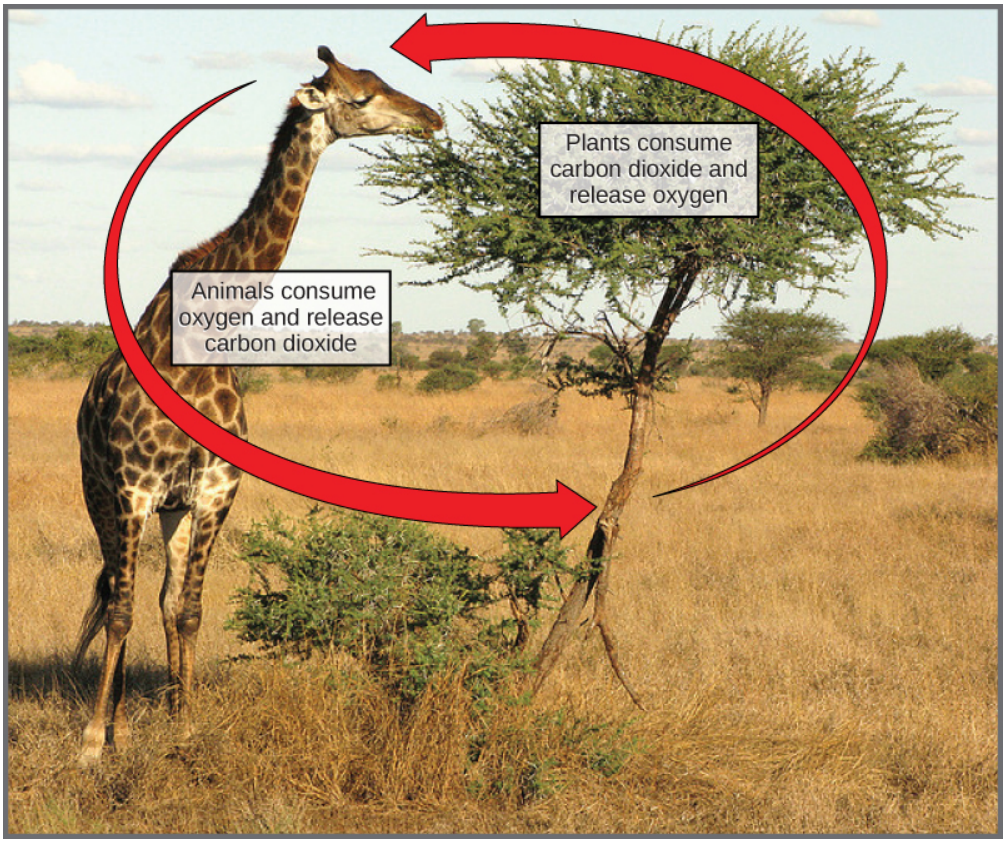


Figure 5.18 In the carbon cycle, the reactions of photosynthesis and cellular respiration share reciprocal reactants and products. (credit: modification of work by Stuart Bassil)

Section Summary

Using the energy carriers formed in the first stage of photosynthesis, the Calvin cycle reactions fix CO_2 from the environment to build carbohydrate molecules. An enzyme, RuBisCO, catalyzes the fixation reaction, by combining CO_2 with RuBP. The resulting six-carbon compound is broken down into two three-carbon compounds, and the energy in ATP and NADPH is used to convert these molecules into G3P. One of the three-carbon molecules of G3P leaves the cycle to become a part of a carbohydrate molecule. The remaining G3P molecules stay in the cycle to be formed back into RuBP, which is ready to react with more CO_2. Photosynthesis forms a balanced energy cycle with the process of cellular

respiration. Plants are capable of both photosynthesis and cellular respiration, since they contain both chloroplasts and mitochondria.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4485#h5p-36

Glossary

Calvin cycle: the reactions of photosynthesis that use the energy stored by the light-dependent reactions to form glucose and other carbohydrate molecules

carbon fixation: the process of converting inorganic CO_2 gas into organic compounds

UNIT 2: CELL DIVISION AND GENETICS

Unit 2: Cell Division and Genetics includes the following chapters:

Chapter 6: Introduction to Reproduction at the Cellular Level

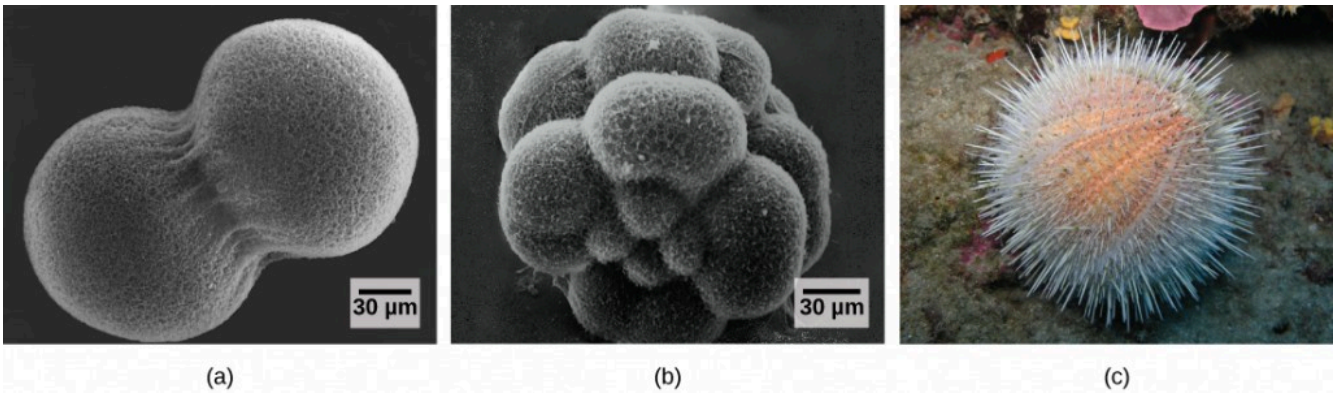


Figure 6.1 A sea urchin begins life as a single cell that (a) divides to form two cells, visible by scanning electron microscopy. After four rounds of cell division, (b) there are 16 cells, as seen in this SEM image. After many rounds of cell division, the individual develops into a complex, multicellular organism, as seen in this (c) mature sea urchin. (credit a: modification of work by Evelyn Spiegel, Louisa Howard; credit b: modification of work by Evelyn Spiegel, Louisa Howard; credit c: modification of work by Marco Busdraghi; scale-bar data from Matt Russell)

The individual sexually reproducing organism—including humans—begins life as a fertilized egg, 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 was the ancestor of every other cell in the body. Once a human individual 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 in most cases, the *maintenance* and *repair* of cells and tissues. Single-celled organisms use cell division as their method of reproduction.

Search for Key Points in Chapter 6

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4490#h5p-37

6.1 The Genome

Learning Objectives

By the end of this section, you will be able to:

- Describe the prokaryotic and eukaryotic genome
- Distinguish between chromosomes, genes, and traits

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 in the life of a cell from the division of a single parent cell to produce two new daughter cells, to the subsequent division of those daughter cells. The mechanisms involved in the cell cycle are highly conserved across eukaryotes. Organisms as diverse as protists, plants, and animals employ similar steps.

Genomic DNA

Before discussing the steps a cell undertakes to replicate, a deeper understanding of the structure and function of a cell's genetic information is necessary. A cell's complete complement of DNA 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. 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.

In **eukaryotes**, the genome comprises **several double-stranded, linear DNA molecules** (Figure 6.2) bound with proteins to form complexes called chromosomes. Each species of eukaryote has a characteristic number of chromosomes in the nuclei of its cells. Human body cells (somatic cells) have **46 chromosomes**. A 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 **23 chromosomes are called gametes**, or sex cells; these eggs and sperm are designated ***n*, or haploid**.

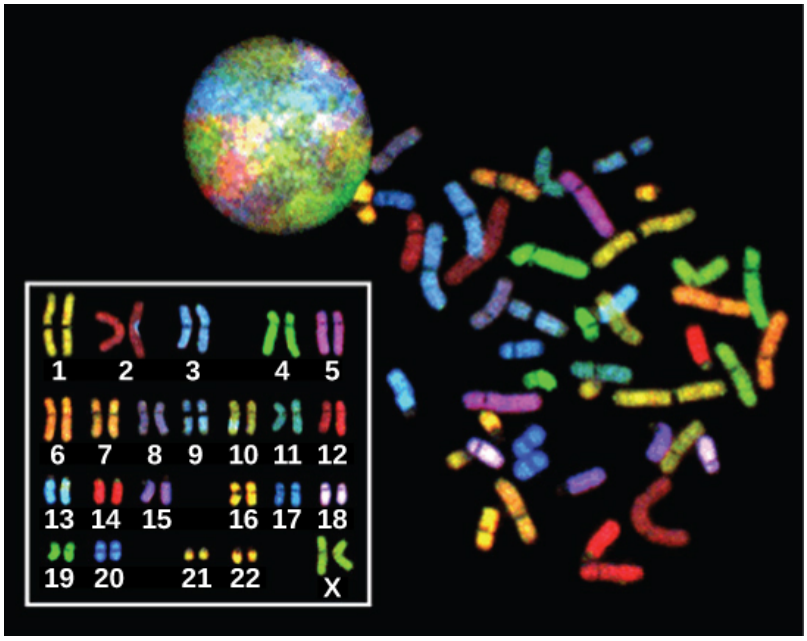


Figure 6.2 There are 23 pairs of homologous chromosomes in a female human somatic cell. These chromosomes are viewed within the nucleus (top), removed from a cell in mitosis (right), and arranged according to length (left) in an arrangement called a karyotype. In this image, the chromosomes were exposed to fluorescent stains to distinguish them. (credit: "718 Bot"/Wikimedia Commons, National Human Genome Research)

The matched pairs of chromosomes in a diploid organism are called homologous 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 different forms of a characteristic. For example, the shape of earlobes is a characteristic with traits of free or attached.

Each copy of the homologous pair of chromosomes originates from a different parent; therefore, the copies of each of the genes themselves may not be identical. The variation of individuals within a species is caused by the specific combination of the genes inherited from both parents. For example, there are three possible gene sequences on the human chromosome that codes 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, one on each homologous chromosome (for example, AA, BB, or OO), or two different sequences, such as AB.

Minor variations in traits such as those for blood type, eye color, and height contribute to the natural variation found within a species. The sex chromosomes, X and Y, are the single exception to the rule of homologous chromosomes; other than a small amount of homology that is necessary to reliably produce gametes, the genes found on the X and Y chromosomes are not the same.

Section Summary

Prokaryotes have a single loop chromosome, whereas eukaryotes have multiple, linear chromosomes surrounded by a nuclear membrane. Human somatic cells have 46 chromosomes consisting of two sets of 22 homologous chromosomes and a pair of nonhomologous sex chromosomes. This is the 2*n*, or diploid, state. Human gametes have 23 chromosomes or one complete set of chromosomes. This is the *n*, or haploid, state. Genes are segments of DNA that code for a specific protein or RNA molecule. An organism's traits are determined in large part by the genes inherited from each parent, but also by the environment that they experience. Genes are expressed as characteristics of the organism and each characteristic may have different variants called traits that are caused by differences in the DNA sequence for a gene.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here: https://opentextbc.ca/biology/?p=4493#h5p-38

Glossary

diploid: describes a cell, nucleus, or organism containing two sets of chromosomes (2*n*)

gamete: a haploid reproductive cell or sex cell (sperm or egg)

gene: the physical and functional unit of heredity; a sequence of DNA that codes for a specific peptide or RNA molecule

genome: the entire genetic complement (DNA) of an organism

haploid: describes a cell, nucleus, or organism containing one set of chromosomes (*n*)

homologous chromosomes: chromosomes of the same length with genes in the same location; diploid organisms have pairs of homologous chromosomes, and the members of each pair come from different parents

locus: the position of a gene on a chromosome

6.2 The Cell Cycle

Learning Objectives

By the end of this section, you will be able to:

- Describe the three stages of interphase
- Discuss the behavior of chromosomes during mitosis and how the cytoplasmic content divides during cytokinesis
- Define the quiescent G_0 phase
- Explain how the three internal control checkpoints occur at the end of G_1, at the G_2–M transition, and during metaphase

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 6.3). 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.

Watch this video about the cell cycle: https://www.youtube.com/watch?v=Wy3N5NCZBHQ

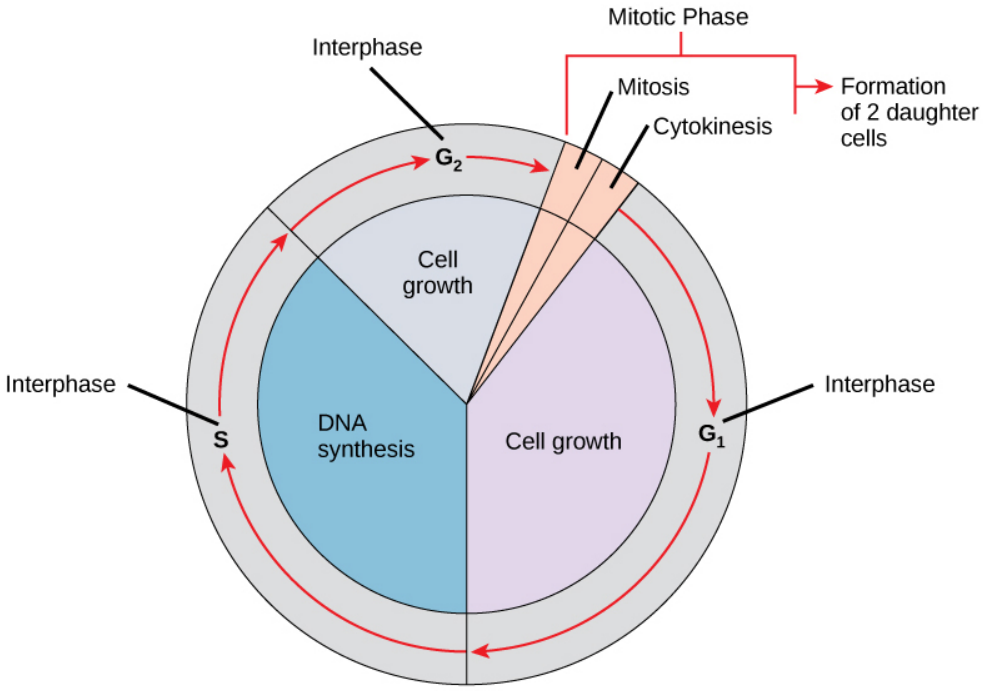


Figure 6.3 A cell moves through a series of phases in an orderly manner. During interphase, G1 involves cell growth and protein synthesis, the S phase involves DNA replication and the replication of the centrosome, and G2 involves further growth and protein synthesis. The mitotic phase follows interphase. 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. 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 dismantled to provide resources 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, prometaphase, metaphase, anaphase, and telophase—that result in the division of the cell nucleus (Figure 6.4).

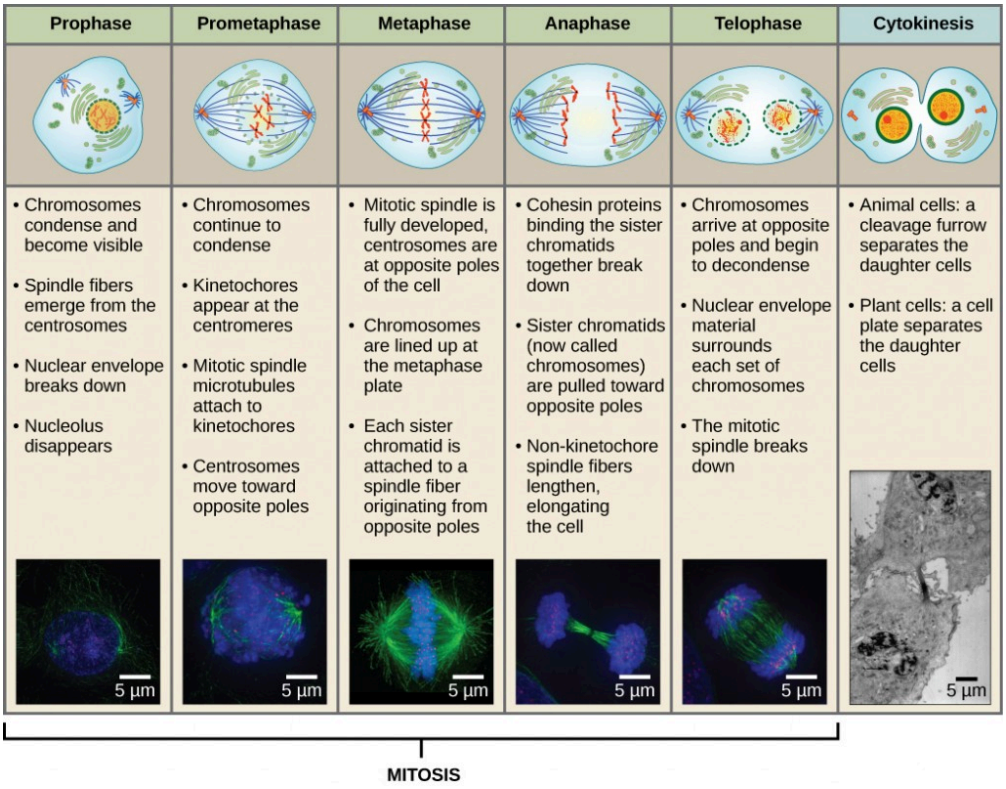


Figure 6.4 Animal cell mitosis is divided into five stages—prophase, prometaphase, metaphase, anaphase, and telophase—visualized here by light microscopy with fluorescence. Mitosis is usually accompanied by cytokinesis, shown here by a transmission electron microscope. (credit "diagrams": modification 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)

Which of the following is the correct order of events in mitosis?

1. Sister chromatids line up at the metaphase plate. The kinetochore becomes attached to the mitotic spindle. The nucleus re-forms and the cell divides. The sister chromatids separate.
2. The kinetochore becomes attached to the mitotic spindle. The sister chromatids separate. Sister chromatids line up at the metaphase plate. The nucleus re-forms and the cell divides.
3. The kinetochore becomes attached to metaphase plate. Sister chromatids line up at the metaphase plate. The kinetochore breaks down and the sister chromatids separate. The nucleus re-forms and the cell divides.
4. The kinetochore becomes attached to the mitotic spindle. Sister chromatids line up at the metaphase plate. The kinetochore breaks apart and the sister chromatids separate. The nucleus re-forms and the cell divides.

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.

During prometaphase, many processes that were begun in prophase continue to advance and culminate in the formation of a connection between the chromosomes and cytoskeleton. The remnants of the nuclear envelope disappear. The mitotic spindle continues to develop as more microtubules assemble and stretch across the length of the former nuclear area. Chromosomes become more condensed and visually discrete. 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.

Concept in Action

This page of movies illustrates different aspects of mitosis. Watch the movie entitled “DIC microscopy of cell division in a newt lung cell” and identify the phases of mitosis.

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 6.5).

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 6.5).

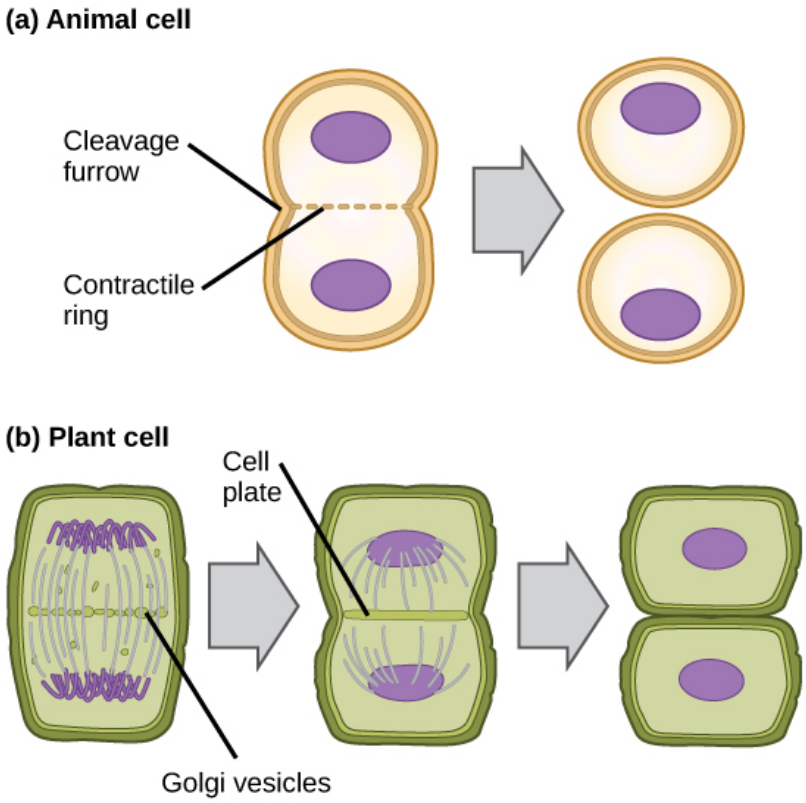


Figure 6.5 In part (a), a cleavage furrow forms at the former metaphase plate in the animal cell. The plasma membrane is drawn in by a ring of actin fibers contracting just inside the membrane. The cleavage furrow deepens until the cells are pinched in two. In part (b), Golgi vesicles coalesce at the 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 6.6).

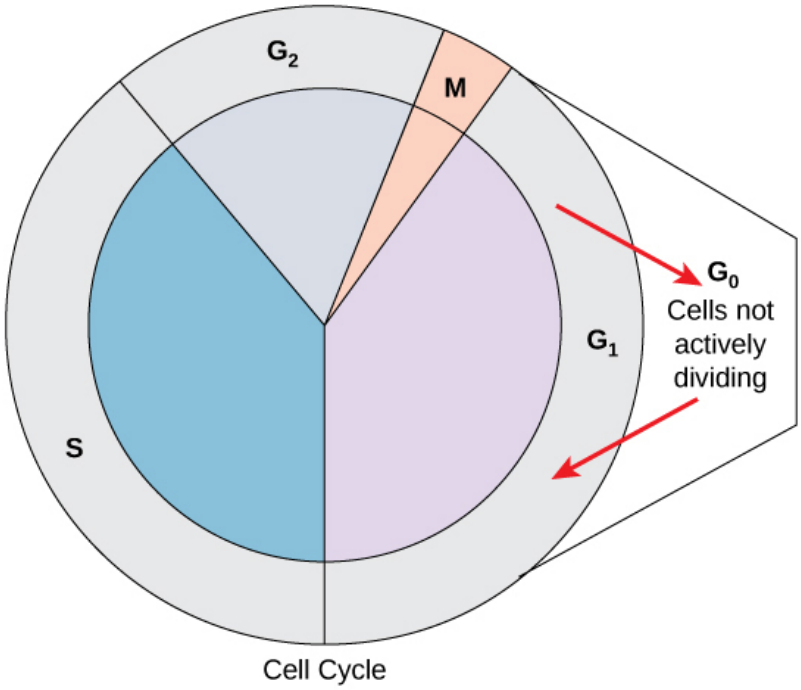


Figure 6.6 Cells that are not actively preparing to divide enter an alternate phase called G0. In some cases, this is a temporary condition until triggered to enter G1. In other cases, the cell will remain in G0 permanently.

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 6.7).

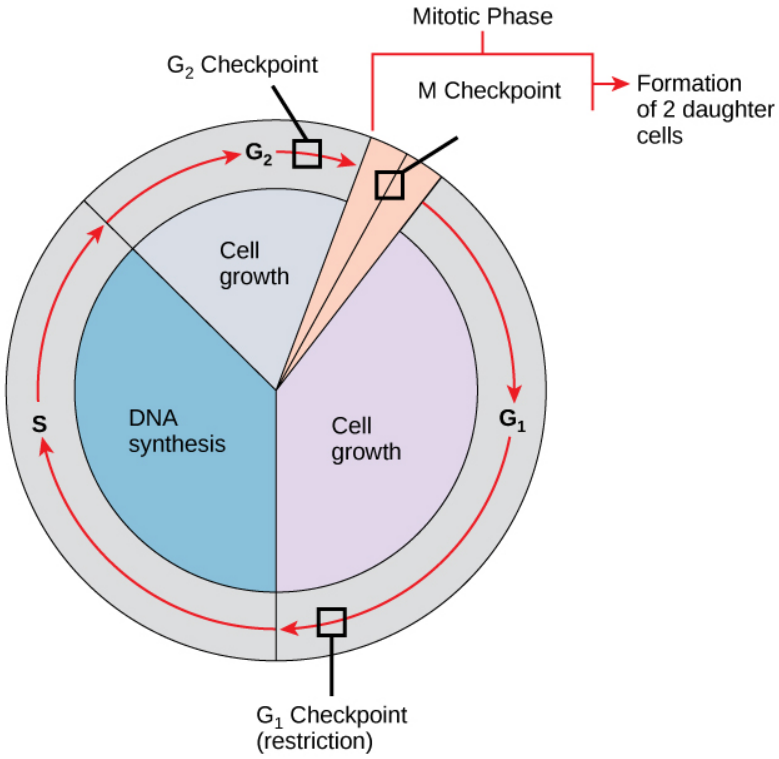


Figure 6.7 The cell cycle is controlled at three checkpoints. Integrity of the DNA is assessed at the G1 checkpoint. Proper chromosome duplication is assessed at the G2 checkpoint. Attachment of each kinetochore to a spindle fiber is assessed at the M checkpoint.

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.

Concept in Action

Watch what occurs at the G_1, G_2, and M checkpoints by visiting this animation of the cell cycle.

Section Summary

The cell cycle is an orderly sequence of events. Cells on the path to cell division proceed through a series of precisely timed and carefully regulated stages. In eukaryotes, the cell cycle consists of a long preparatory period, called interphase. Interphase is divided into G_1, S, and G_2 phases. Mitosis consists of five stages: prophase, prometaphase, metaphase, anaphase, and telophase. Mitosis is usually accompanied by cytokinesis, during which the cytoplasmic components of the daughter cells are separated either by an actin ring (animal cells) or by cell plate formation (plant cells).

Each step of the cell cycle is monitored by internal controls called checkpoints. There are three major checkpoints in the cell cycle: one near the end of G_1, a second at the G_2–M transition, and the third during metaphase.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4502#h5p-39

Glossary

anaphase: the stage of mitosis during which sister chromatids are separated from each other

cell cycle: the ordered sequence of events that a cell passes through between one cell division and the next

cell cycle checkpoints: mechanisms that monitor the preparedness of a eukaryotic cell to advance through the various cell cycle stages

cell plate: a structure formed during plant-cell cytokinesis by Golgi vesicles fusing at the metaphase plate; will ultimately lead to formation of a cell wall to separate the two daughter cells

centriole: a paired rod-like structure constructed of microtubules at the center of each animal cell centrosome

cleavage furrow: a constriction formed by the actin ring during animal-cell cytokinesis that leads to cytoplasmic division

cytokinesis: the division of the cytoplasm following mitosis to form two daughter cells

G_0 phase: a cell-cycle phase distinct from the G_1 phase of interphase; a cell in G_0 is not preparing to divide

G_1 phase: (also, first gap) a cell-cycle phase; first phase of interphase centered on cell growth during mitosis

G_2 phase: (also, second gap) a cell-cycle phase; third phase of interphase where the cell undergoes the final preparations for mitosis

interphase: the period of the cell cycle leading up to mitosis; includes G_1, S, and G_2 phases; the interim between two consecutive cell divisions

kinetochore: a protein structure in the centromere of each sister chromatid that attracts and binds spindle microtubules during prometaphase

metaphase plate: the equatorial plane midway between two poles of a cell where the chromosomes align during metaphase

metaphase: the stage of mitosis during which chromosomes are lined up at the metaphase plate

mitosis: the period of the cell cycle at which the duplicated chromosomes are separated into identical nuclei; includes prophase, prometaphase, metaphase, anaphase, and telophase

mitotic phase: the period of the cell cycle when duplicated chromosomes are distributed into two nuclei and the cytoplasmic contents are divided; includes mitosis and cytokinesis

mitotic spindle: the microtubule apparatus that orchestrates the movement of chromosomes during mitosis

prometaphase: the stage of mitosis during which mitotic spindle fibers attach to kinetochores

prophase: the stage of mitosis during which chromosomes condense and the mitotic spindle begins to form

quiescent: describes a cell that is performing normal cell functions and has not initiated preparations for cell division

S phase: the second, or synthesis phase, of interphase during which DNA replication occurs

telophase: the stage of mitosis during which chromosomes arrive at opposite poles, decondense, and are surrounded by new nuclear envelopes

6.3 Cancer and the Cell Cycle

Learning Objectives

By the end of this section, you will be able to:

- Explain how cancer is caused by uncontrolled cell division
- Understand how proto-oncogenes are normal cell genes that, when mutated, become oncogenes
- Describe how tumor suppressors function to stop the cell cycle until certain events are completed
- Explain how mutant tumor suppressors cause 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 cell-cycle control, 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 tumor can result.

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 6.8). 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.

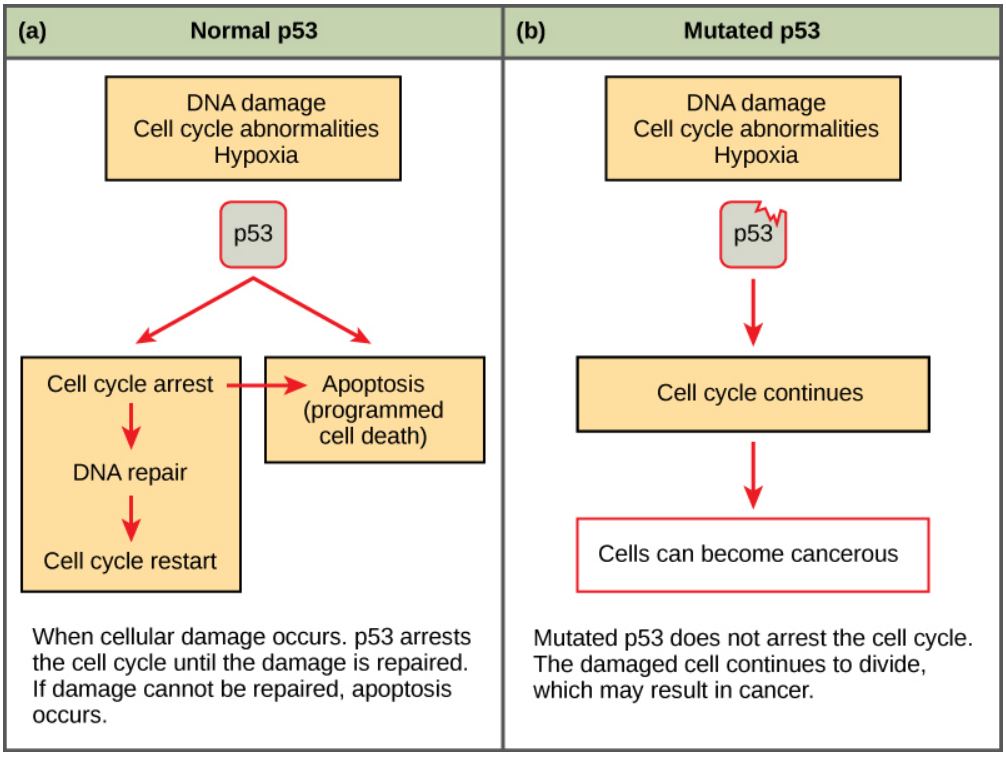
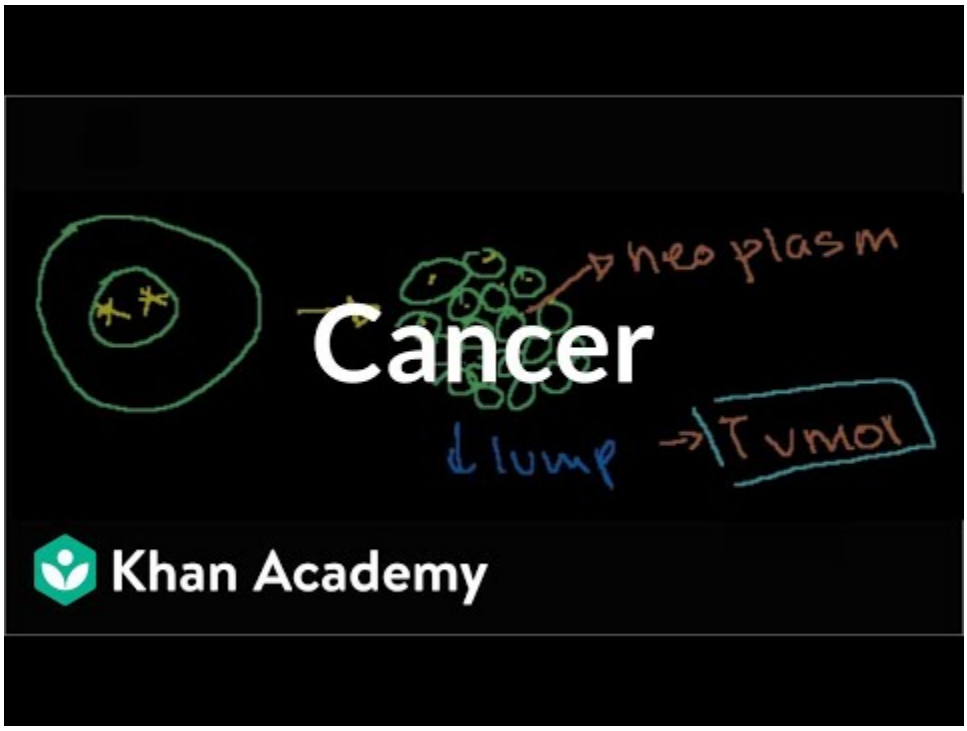


Figure 6.8 (a) The role of p53 is to monitor DNA. If damage is detected, p53 triggers repair mechanisms. If repairs are unsuccessful, p53 signals apoptosis. (b) A cell with an abnormal p53 protein cannot repair damaged DNA and cannot signal apoptosis. Cells with abnormal p53 can become cancerous. (credit: modification of work by Thierry Soussi)

Concept in Action



A YouTube element has been excluded from this version of the text. You can view it online here: https://opentextbc.ca/biology/?p=4506

Go to this website to watch an animation of how cancer results from errors in the cell cycle.

Section Summary

Cancer is the result of unchecked cell division caused by a breakdown of the mechanisms regulating the cell cycle. The loss of control begins with a change in the DNA sequence of a gene that codes for one of the regulatory molecules. Faulty instructions lead to a protein that does not function as it should. Any disruption of the monitoring system can allow other mistakes to be passed on to the daughter cells. Each successive cell division will give rise to daughter cells with even more accumulated damage. Eventually, all checkpoints become nonfunctional, and rapidly reproducing cells crowd out normal cells, resulting in tumorous growth.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4506#h5p-40

Glossary

oncogene: a mutated version of a proto-oncogene, which allows for uncontrolled progression of the cell cycle, or uncontrolled cell reproduction

proto-oncogene: a normal gene that controls cell division by regulating the cell cycle that becomes an oncogene if it is mutated

tumor suppressor gene: a gene that codes for regulator proteins that prevent the cell from undergoing uncontrolled division

6.4 Prokaryotic Cell Division

Learning Objectives

By the end of this section, you will be able to:

- Describe the process of binary fission in prokaryotes
- Explain how FtsZ and tubulin proteins are examples of homology

Prokaryotes such as bacteria propagate by binary fission. For unicellular organisms, cell division is the only method to produce new individuals. In both prokaryotic and eukaryotic cells, the outcome of cell reproduction is a pair of daughter cells that are genetically identical to the parent cell. In unicellular organisms, daughter cells are individuals.

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 6.9). 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.

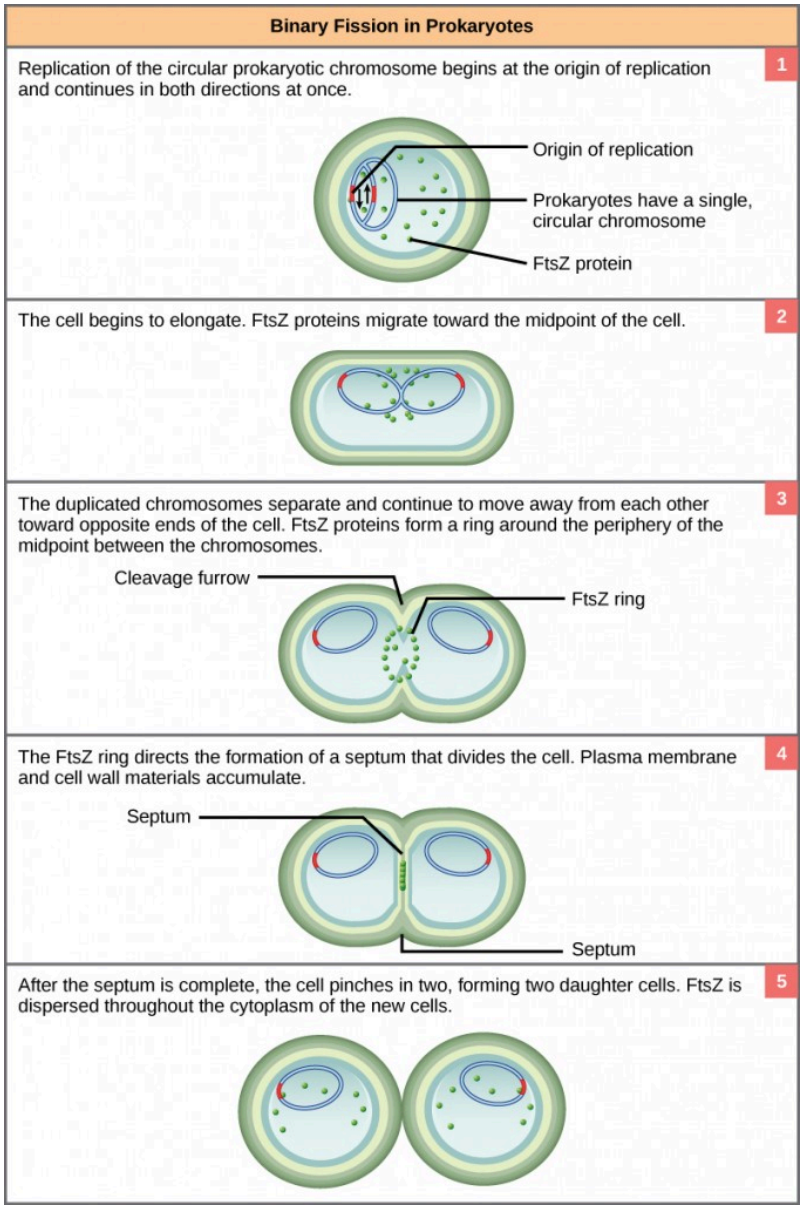


Figure 6.9 The binary fission of a bacterium is outlined in five steps. (credit: modification of work by "Mcstrother"/Wikimedia Commons)

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.

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**
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.
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.

Section Summary

In both prokaryotic and eukaryotic cell division, the genomic DNA is replicated and each copy is allocated into a daughter cell. The cytoplasmic contents are also divided evenly to the new cells. However, there are many differences between prokaryotic and eukaryotic cell division. Bacteria have a single, circular DNA chromosome and no nucleus. Therefore, mitosis is not necessary in bacterial cell division. Bacterial cytokinesis is directed by a ring composed of a protein called FtsZ. Ingrowth of membrane and cell-wall material from the periphery of the cells results in a septum that eventually forms the separate cell walls of the daughter cells.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4509#h5p-41

Glossary

binary fission: the process of prokaryotic cell division

FtsZ: a tubulin-like protein component of the prokaryotic cytoskeleton that is important in prokaryotic cytokinesis (name origin: **F**ilamenting **t**emperature-**s**ensitive mutant **Z**)

origin: the region of the prokaryotic chromosome at which replication begins

septum: a wall formed between bacterial daughter cells as a precursor to cell separation

Chapter 7: Introduction to the Cellular Basis of Inheritance

Figure 7.1 Each of us, like these other large multicellular organisms, begins life as a fertilized egg. After trillions of cell divisions, each of us develops into a complex, multicellular organism. (credit a: modification of work by Frank Wouters; credit b: modification of work by Ken Cole, USGS; credit c: modification of work by Martin Pettitt)

The ability to reproduce *in kind* is a basic characteristic of all living things. *In kind* means that the offspring of any organism closely resembles its parent or parents. Hippopotamuses give birth to hippopotamus calves; Monterey pine trees produce seeds from which Monterey pine seedlings emerge; and adult flamingos lay eggs that hatch into flamingo chicks. ***In kind* does not generally mean *exactly the same***. While many single-celled organisms and a few multicellular organisms can produce genetically identical clones of themselves through mitotic cell division, many single-celled organisms and most multicellular organisms reproduce regularly using another method.

Sexual reproduction is the production by parents of haploid cells and the fusion of a haploid cell from each parent to form a single, unique diploid cell. In multicellular organisms, the new diploid cell will then undergo mitotic cell divisions to develop into an adult organism. A type of cell division called meiosis leads to the haploid cells that are part of the sexual reproductive cycle. Sexual reproduction, specifically meiosis and fertilization, introduces variation into offspring that may account for the evolutionary success of sexual reproduction. The vast majority of eukaryotic organisms can or must employ some form of meiosis and fertilization to reproduce.

Search for Key Points in Chapter 7

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4512#h5p-42

7.1 Sexual Reproduction

Learning Objectives

By the end of this section, you will be able to:

- Explain that variation among offspring is a potential evolutionary advantage resulting from sexual reproduction
- Describe the three different life-cycle strategies among sexual multicellular organisms and their commonalities
- Understand why you could never create a gamete that would be identical to either of the gametes that made yo

An interactive H5P element has been excluded from this version of the text. You can view it online here: *https://opentextbc.ca/biology/?p=4515#h5p-43*

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.

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. 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 7.2 **a**).

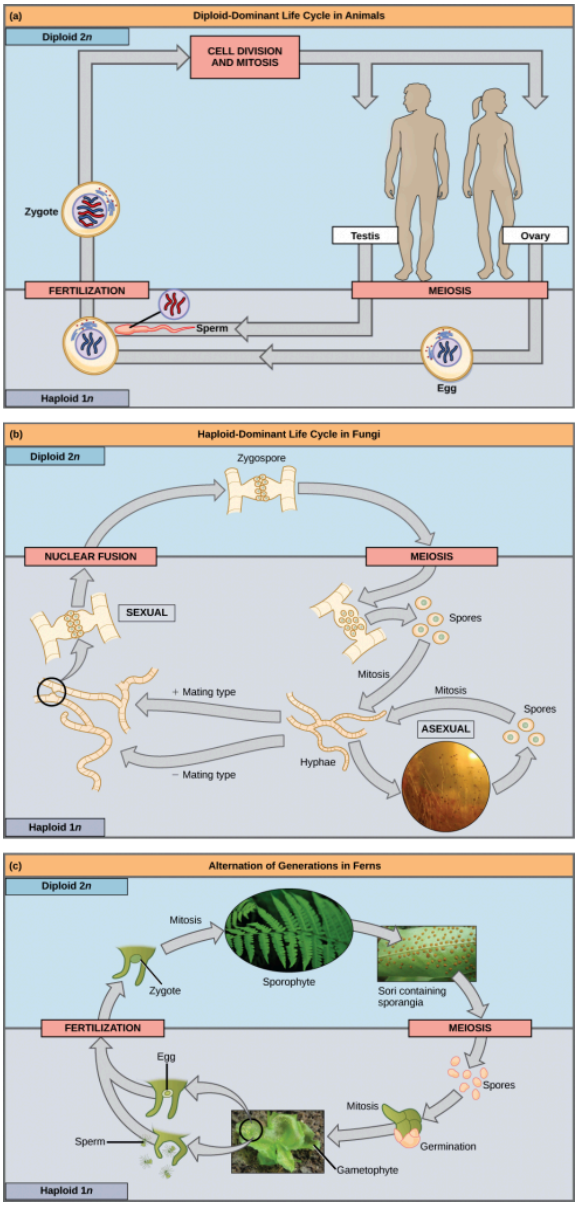


Figure 7.2 (a) In animals, sexually reproducing adults form haploid gametes from diploid germ cells. (b) Fungi, such as black bread mold (Rhizopus nigricans), have haploid-dominant life cycles. (c) Plants have a life cycle that alternates between a multicellular haploid organism and a multicellular diploid organism. (credit c "fern": modification of work by Cory Zanker; credit c "gametophyte": modification of work by "Vlmastra"/Wikimedia Commons)

If a mutation occurs so that a fungus is no longer able to produce a minus mating type, will it still be able to reproduce?

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 7.2 **b**).

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 7. 2 **c**).

Section Summary

Nearly all eukaryotes undergo sexual reproduction. The variation introduced into the reproductive cells by meiosis appears to be one of the advantages of sexual reproduction that has made it so successful. Meiosis and fertilization alternate in sexual life cycles. The process of **meiosis produces genetically unique reproductive cells called gametes**, which have half the number of chromosomes as the parent cell. Fertilization, the fusion of haploid gametes from two individuals, restores the diploid condition. Thus, sexually reproducing organisms alternate between haploid and diploid stages. However, the ways in which reproductive cells are produced and the timing between meiosis and fertilization vary greatly. There are three main categories of life cycles: diploid-dominant, demonstrated by most animals; haploid-dominant, demonstrated by all fungi and some algae; and alternation of generations, demonstrated by plants and some algae.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4515#h5p-44

Glossary

alternation of generations: a life-cycle type in which the diploid and haploid stages alternate

diploid-dominant: a life-cycle type in which the multicellular diploid stage is prevalent

haploid-dominant: a life-cycle type in which the multicellular haploid stage is prevalent

gametophyte: a multicellular haploid life-cycle stage that produces gametes

germ cell: a specialized cell that produces gametes, such as eggs or sperm

life cycle: the sequence of events in the development of an organism and the production of cells that produce offspring

meiosis: a nuclear division process that results in four haploid cells

sporophyte: a multicellular diploid life-cycle stage that produces spores

Footnotes

1 Leigh Van Valen, “A new evolutionary law,” Evolutionary Theory 1 (1973): 1–30.

7.2 Meiosis

Learning Objectives

By the end of this section, you will be able to:

- Describe the behavior of chromosomes during meiosis
- Describe cellular events during meiosis
- Explain the differences between meiosis and mitosis
- Explain the mechanisms within meiosis that generate genetic variation among the products of meiosis

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 cells contain one set of chromosomes. Cells containing two sets of chromosomes are called diploid. 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. 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.

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 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, prometaphase 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, prometaphase II, and so on.

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

Early 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**. This process is revealed visually after the exchange as chiasmata (singular = *chiasma*) (Figure 7.3).

As prophase I progresses, the close association between homologous chromosomes begins to break down, and the chromosomes continue to condense, although the homologous chromosomes remain attached to each other at chiasmata. The number of chiasmata varies with the species and the length of the chromosome. At the end of prophase I, the pairs are held together only at chiasmata (Figure 7.3) and are called tetrads because the four sister chromatids of each pair of homologous chromosomes are now visible.

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.

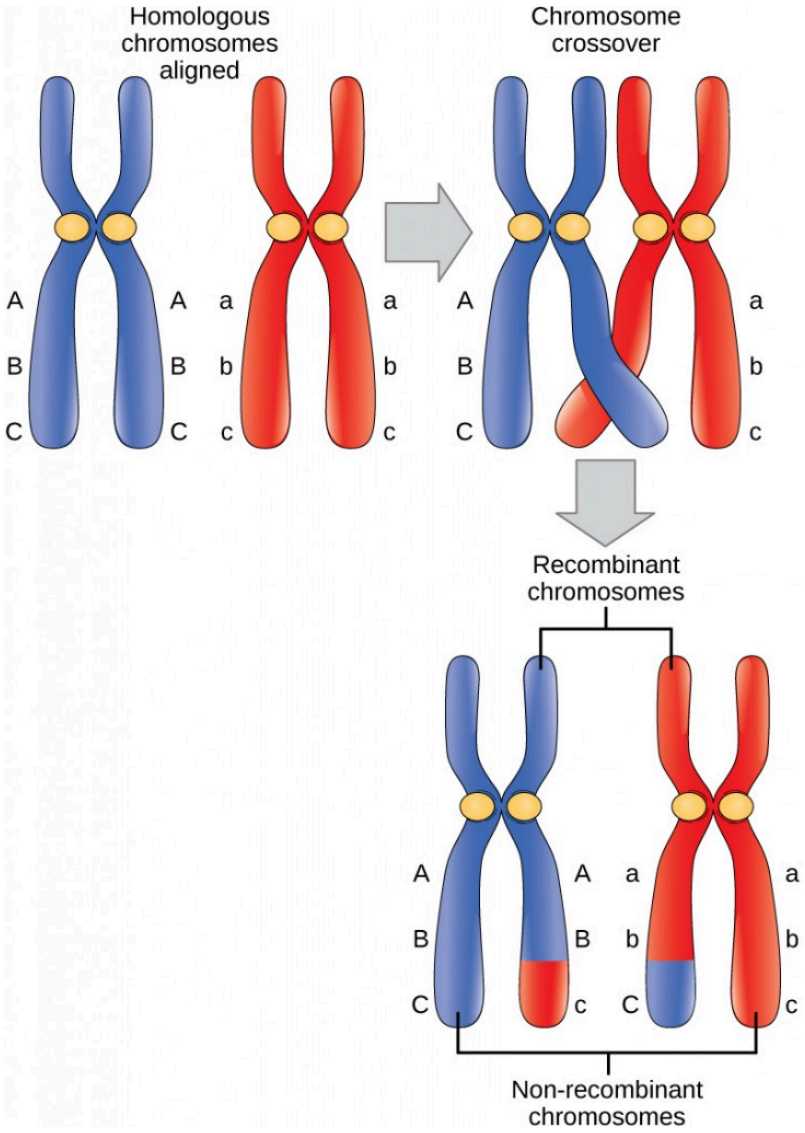


Figure 7.3 In this illustration of the effects of crossing over, the blue chromosome came from the individual's father and the red chromosome came from the individual's mother. Crossover occurs between non-sister chromatids of homologous chromosomes. The result is an exchange of genetic material between homologous chromosomes. The chromosomes that have a mixture of maternal and paternal sequence are called recombinant and the chromosomes that are completely paternal or maternal are called non-recombinant.

The key event in prometaphase 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. At the end of prometaphase 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. The homologous chromosomes are still held together at chiasmata. In addition, the nuclear membrane has broken down entirely.

During 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 7.4).

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.

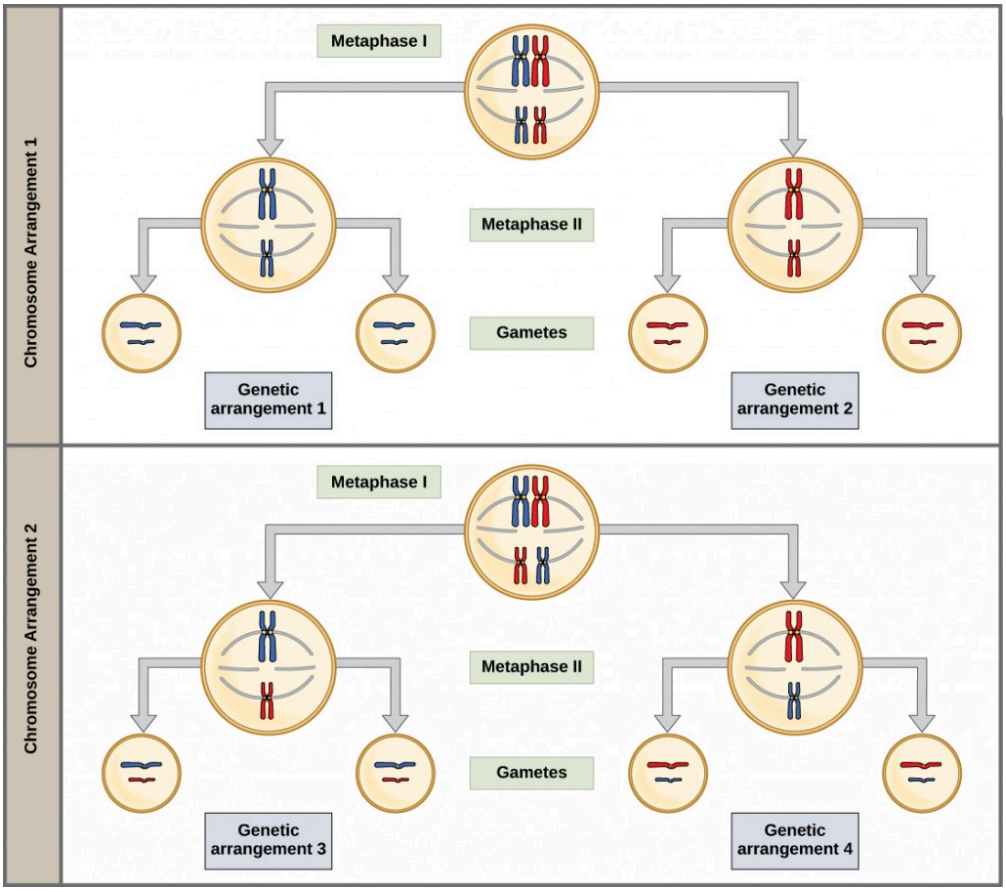


Figure 7.4 To demonstrate random, independent assortment at metaphase I, consider a cell with n = 2. In this case, there are two possible arrangements at the equatorial plane in metaphase I, as shown in the upper cell of each panel. These two possible orientations lead to the production of genetically different gametes. With more chromosomes, the number of possible arrangements increases dramatically.

In anaphase I, the spindle fibers pull the linked chromosomes apart. The sister chromatids remain tightly bound together at the centromere. It is the chiasma connections that are broken in anaphase I as the fibers attached to the fused kinetochores pull the homologous chromosomes apart.

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.

Concept in Action

Review the process of meiosis, observing how chromosomes align and migrate, at this site.

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. 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 prometaphase 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.

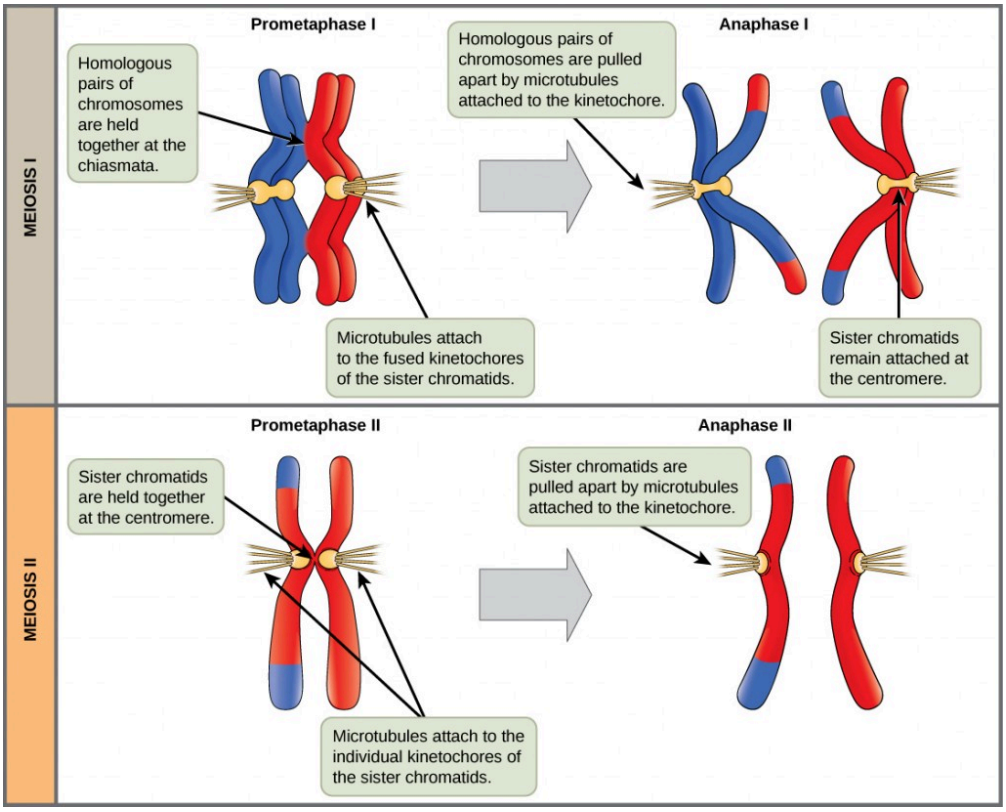


Figure 7.5 In prometaphase I, microtubules attach to the fused kinetochores of homologous chromosomes. In anaphase I, the homologous chromosomes are separated. In prometaphase II, microtubules attach to individual kinetochores of sister chromatids. In anaphase II, the sister chromatids are separated.

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.

Comparing Meiosis and Mitosis

Mitosis and meiosis, which are both forms of division of the nucleus in eukaryotic cells, share some similarities, but also exhibit distinct differences that lead to their very different outcomes. Mitosis is a single nuclear division that results in two nuclei, usually partitioned into two new cells. The nuclei resulting from a mitotic division are genetically identical to the original. They have the same number of sets of chromosomes: one in the case of haploid cells, and two in the case of diploid cells. On the other hand, meiosis is two nuclear divisions that result in four nuclei, usually partitioned into four new cells. The nuclei resulting from meiosis are never genetically identical, and they contain one chromosome set only—this is half the number of the original cell, which was diploid.

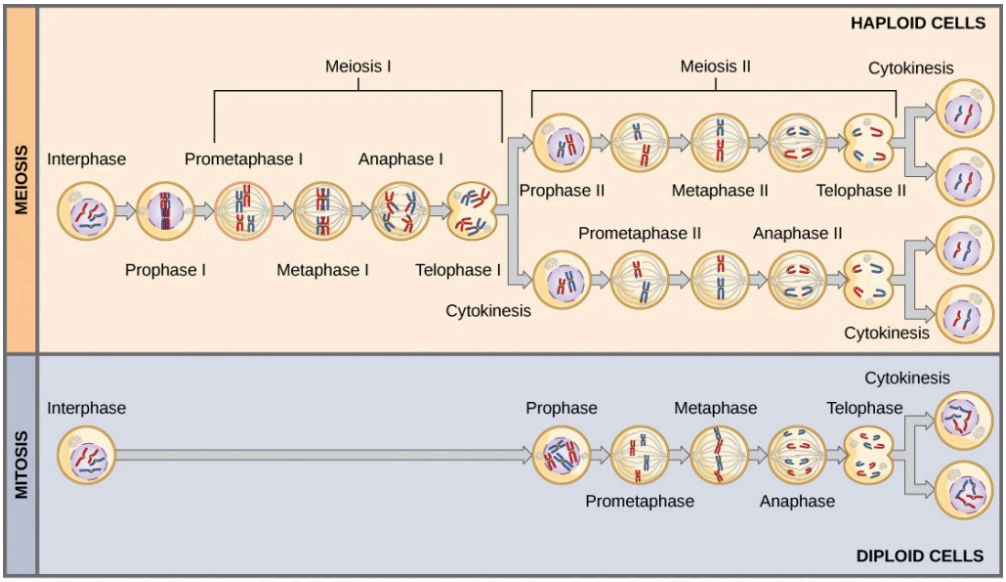
The differences in the outcomes of meiosis and mitosis occur because of differences in the behavior of the chromosomes during each process. Most of these differences in the processes occur in meiosis I,

which is a very different nuclear division than mitosis. In meiosis I, the homologous chromosome pairs become associated with each other, are bound together, experience chiasmata and crossover between sister chromatids, and line up along the metaphase plate in tetrads with spindle fibers from opposite spindle poles attached to each kinetochore of a homolog in a tetrad. All of these events occur only in meiosis I, never in mitosis.

Homologous chromosomes move to opposite poles during meiosis I so the number of sets of chromosomes in each nucleus-to-be is reduced from two to one. For this reason, meiosis I is referred to as a reduction division. There is no such reduction in ploidy level in mitosis.

Meiosis II is much more analogous to a mitotic division. In this case, duplicated chromosomes (only one set of them) line up at the center of the cell with divided kinetochores attached to spindle fibers from opposite poles. During anaphase II, as in mitotic anaphase, the kinetochores divide and one sister chromatid is pulled to one pole and the other sister chromatid is pulled to the other pole. If it were not for the fact that there had been crossovers, the two products of each meiosis II division would be identical as in mitosis; instead, they are different because there has always been at least one crossover per chromosome. Meiosis II is not a reduction division because, although there are fewer copies of the genome in the resulting cells, there is still one set of chromosomes, as there was at the end of meiosis I.

Cells produced by mitosis will function in different parts of the body as a part of growth or replacing dead or damaged cells. They may even be involved in asexual reproduction in some organisms. Cells produced by meiosis in a diploid-dominant organism such as an animal will only participate in sexual reproduction.



						OUTCOME
PROCESS	DNA synthesis	Synapsis of homologous chromosomes	Crossover	Homologous chromosomes line up at metaphase plate	Sister chromatids line up at metaphase plate	Number and genetic composition of daughter cells
MEIOSIS	Occurs in S phase of interphase	During prophase I	During prophase I	During metaphase I	During metaphase II	Four haploid cells at the end of meiosis II
MITOSIS	Occurs in S phase of interphase	Does not occur in mitosis	Does not occur in mitosis	Does not occur in mitosis	During metaphase	Two diploid cells at the end of mitosis

Figure 7.6 Meiosis and mitosis are both preceded by one round of DNA replication; however, meiosis includes two nuclear divisions. The four daughter cells resulting from meiosis are haploid and genetically distinct. The daughter cells resulting from mitosis are diploid and identical to the parent cell.

Concept in Action

For an animation comparing mitosis and meiosis, go to this website.

Section Summary

Sexual reproduction requires that diploid organisms produce haploid cells that can fuse during

fertilization to form diploid offspring. The process that results in haploid cells is called meiosis. Meiosis is a series of events that arrange and separate chromosomes into daughter cells. During the interphase of meiosis, each chromosome is duplicated. In meiosis, there are two rounds of nuclear division resulting in four nuclei and usually four haploid daughter cells, each with half the number of chromosomes as the parent cell. During meiosis, variation in the daughter nuclei is introduced because of crossover in prophase I and random alignment at metaphase I. The cells that are produced by meiosis are genetically unique.

Meiosis and mitosis share similarities, but have distinct outcomes. Mitotic divisions are single nuclear divisions that produce daughter nuclei that are genetically identical and have the same number of chromosome sets as the original cell. Meiotic divisions are two nuclear divisions that produce four daughter nuclei that are genetically different and have one chromosome set rather than the two sets the parent cell had. The main differences between the processes occur in the first division of meiosis. The homologous chromosomes separate into different nuclei during meiosis I causing a reduction of ploidy level. The second division of meiosis is much more similar to a mitotic division.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4523#h5p-45

Glossary

chiasmata: (singular = *chiasma*) the structure that forms at the crossover points after genetic material is exchanged

crossing over: (also, recombination) the exchange of genetic material between homologous chromosomes resulting in chromosomes that incorporate genes from both parents of the organism forming reproductive cells

fertilization: the union of two haploid cells typically from two individual organisms

interkinesis: a period of rest that may occur between meiosis I and meiosis II; there is no replication of DNA during interkinesis

meiosis I: the first round of meiotic cell division; referred to as reduction division because the resulting cells are haploid

meiosis II: the second round of meiotic cell division following meiosis I; sister chromatids are separated from each other, and the result is four unique haploid cells

recombinant: describing something composed of genetic material from two sources, such as a chromosome with both maternal and paternal segments of DNA

reduction division: a nuclear division that produces daughter nuclei each having one-half as many chromosome sets as the parental nucleus; meiosis I is a reduction division

somatic cell: all the cells of a multicellular organism except the gamete-forming cells

synapsis: the formation of a close association between homologous chromosomes during prophase I

tetrad: two duplicated homologous chromosomes (four chromatids) bound together by chiasmata during prophase I

7.3 Errors in Meiosis

Learning Objectives

By the end of this section, you will be able to:

- Explain how nondisjunction leads to disorders in chromosome number
- Describe how errors in chromosome structure occur through inversions and translocations

Inherited disorders can arise when chromosomes behave abnormally during meiosis. Chromosome disorders can be divided into two categories: abnormalities in chromosome number and chromosome structural rearrangements. Because even small segments of chromosomes can span many genes, chromosomal disorders are characteristically dramatic and often fatal.

Disorders in Chromosome Number

The isolation and microscopic observation of chromosomes forms the basis of cytogenetics and is the primary method by which clinicians detect chromosomal abnormalities in humans. A **karyotype** is the number and appearance of chromosomes, including their length, banding pattern, and centromere position. To obtain a view of an individual's karyotype, cytologists photograph the chromosomes and then cut and paste each chromosome into a chart, or karyogram (Figure 7.7).

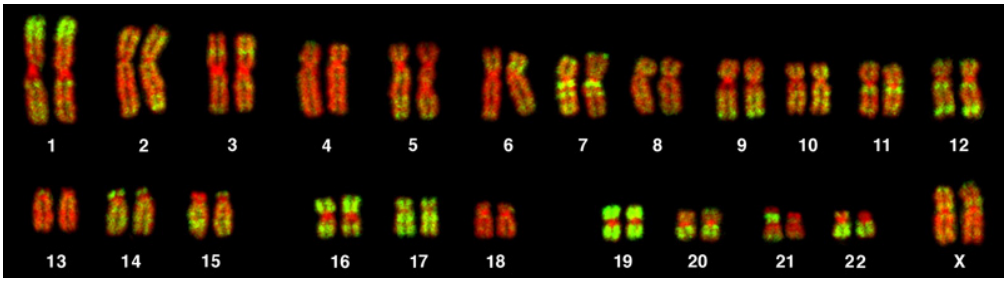


Figure 7.7 This karyogram shows the chromosomes of a female human immune cell during mitosis. (credit: Andreas Bolzer, et al)

Geneticists Use Karyograms to Identify Chromosomal Aberrations

The karyotype is a method by which traits characterized by chromosomal abnormalities can be identified from a single cell. To observe an individual's karyotype, a person's cells (like white blood cells) are first collected from a blood sample or other tissue. In the laboratory, the isolated cells are

stimulated to begin actively dividing. A chemical is then applied to the cells to arrest mitosis during metaphase. The cells are then fixed to a slide.

The geneticist then stains chromosomes with one of several dyes to better visualize the distinct and reproducible banding patterns of each chromosome pair. Following staining, chromosomes are viewed using bright-field microscopy. An experienced cytogeneticist can identify each band. In addition to the banding patterns, chromosomes are further identified on the basis of size and centromere location. To obtain the classic depiction of the karyotype in which homologous pairs of chromosomes are aligned in numerical order from longest to shortest, the geneticist obtains a digital image, identifies each chromosome, and manually arranges the chromosomes into this pattern.

At its most basic, the karyogram may reveal genetic abnormalities in which an individual has too many or too few chromosomes per cell. Examples of this are **Down syndrome**, which is identified by a **third copy of chromosome 21**, and Turner syndrome, which is characterized by the presence of only one X chromosome in women instead of two. Geneticists can also identify large deletions or insertions of DNA. For instance, Jacobsen syndrome, which involves distinctive facial features as well as heart and bleeding defects, is identified by a deletion on chromosome 11. Finally, the karyotype can pinpoint translocations, which occur when a segment of genetic material breaks from one chromosome and reattaches to another chromosome or to a different part of the same chromosome. Translocations are implicated in certain cancers, including chronic myelogenous leukemia.

By observing a karyogram, geneticists can actually visualize the chromosomal composition of an individual to confirm or predict genetic abnormalities in offspring even before birth.

Nondisjunctions, Duplications, and Deletions

Of all the chromosomal disorders, abnormalities in chromosome number are the most easily identifiable from a karyogram. Disorders of chromosome number include the duplication or loss of entire chromosomes, as well as changes in the number of complete sets of chromosomes. They are caused by **nondisjunction**, which occurs when pairs of homologous chromosomes or sister chromatids fail to separate during meiosis. The risk of nondisjunction increases with the age of the parents.

Nondisjunction can occur during either meiosis I or II, with different results (Figure 7.8). If homologous chromosomes fail to separate during meiosis I, the result is two gametes that lack that chromosome and two gametes with two copies of the chromosome. If sister chromatids fail to separate during meiosis II, the result is one gamete that lacks that chromosome, two normal gametes with one copy of the chromosome, and one gamete with two copies of the chromosome.

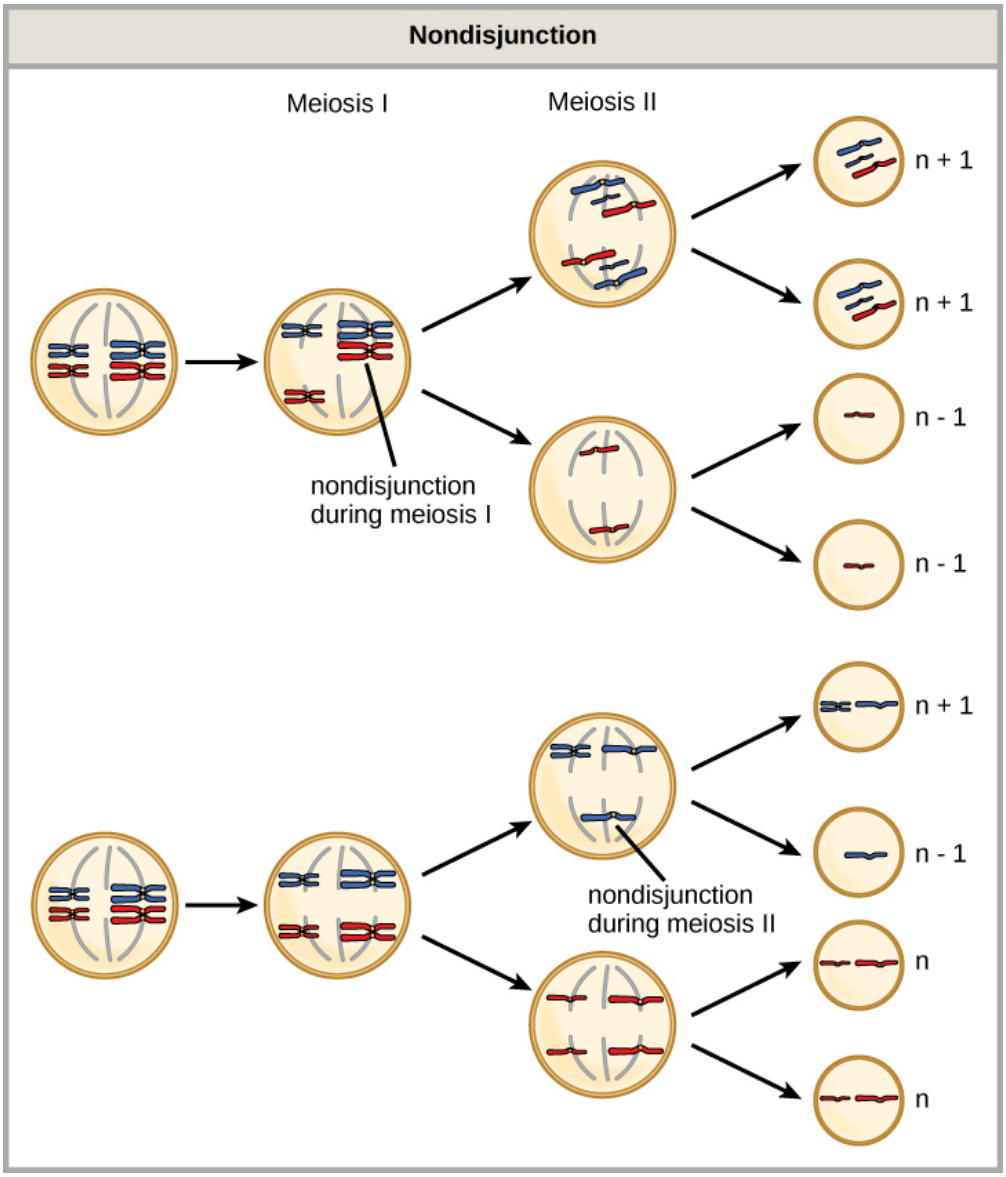


Figure 7.8 Following meiosis, each gamete has one copy of each chromosome. Nondisjunction occurs when homologous chromosomes (meiosis I) or sister chromatids (meiosis II) fail to separate during meiosis.

An individual with the appropriate number of chromosomes for their species is called euploid; in humans, euploidy corresponds to 22 pairs of autosomes and one pair of sex chromosomes. An individual with an error in chromosome number is described as aneuploid, a term that includes monosomy (loss of one chromosome) or trisomy (gain of an extraneous chromosome). Monosomic human zygotes missing any one copy of an autosome invariably fail to develop to birth because they have only one copy of essential genes. Most autosomal trisomies also fail to develop to birth; however, duplications of some of the smaller chromosomes (13, 15, 18, 21, or 22) can result in offspring that survive for several weeks to many years. Trisomic individuals suffer from a different type of genetic imbalance: an excess in gene dose. Cell functions are calibrated to the amount of gene product produced by two copies (doses) of each gene; adding a third copy (dose) disrupts this balance. The

most common trisomy is that of chromosome 21, which leads to Down syndrome. Individuals with this inherited disorder have characteristic physical features and developmental delays in growth and cognition. The incidence of Down syndrome is correlated with maternal age, such that older women are more likely to give birth to children with Down syndrome (Figure 7.9).

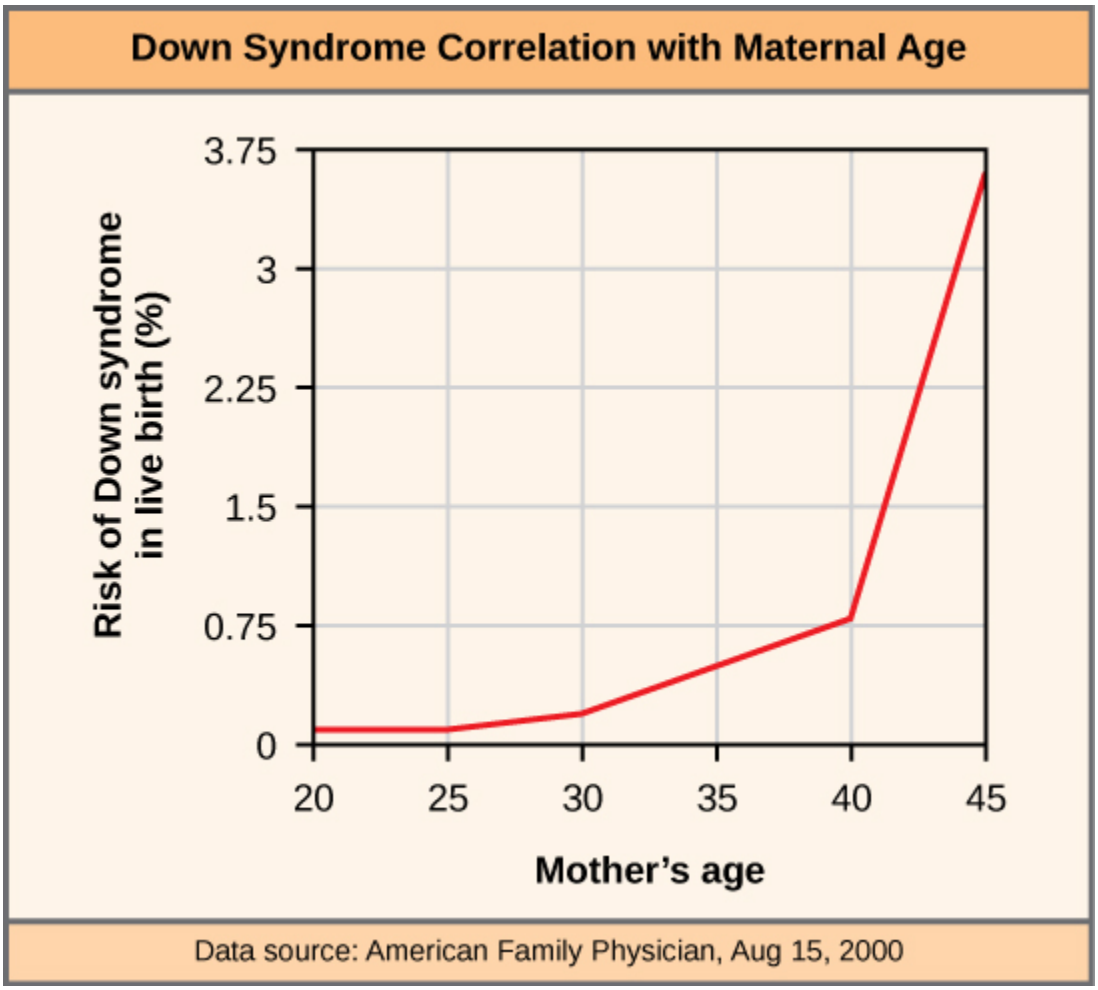


Figure 7.9 The incidence of having a fetus with trisomy 21 increases dramatically with maternal age.

Concept in Action

Visualize the addition of a chromosome that leads to Down syndrome in this video simulation.

Humans display dramatic deleterious effects with autosomal trisomies and monosomies. Therefore, it may seem counterintuitive that human females and males can function normally, despite carrying different numbers of the X chromosome. In part, this occurs because of a process called X inactivation. Early in development, when female mammalian embryos consist of just a few thousand cells, one X chromosome in each cell inactivates by condensing into a structure called a Barr body. The genes on the inactive X chromosome are not expressed. The particular X chromosome (maternally or paternally derived) that is inactivated in each cell is random, but once the inactivation occurs, all cells descended from that cell will have the same inactive X chromosome. By this process, females compensate for their double genetic dose of X chromosome.

In so-called “tortoiseshell” cats, X inactivation is observed as coat-color variegation (Figure 7.10). Females heterozygous for an X-linked coat color gene will express one of two different coat colors over different regions of their body, corresponding to whichever X chromosome is inactivated in the embryonic cell progenitor of that region. When you see a tortoiseshell cat, you will know that it has to be a female.

Figure 7.10 Embryonic inactivation of one of two different X chromosomes encoding different coat colors gives rise to the tortoiseshell phenotype in cats. (credit: Michael Bodega) Photo of a tortoiseshell cat.

In an individual carrying an abnormal number of X chromosomes, cellular mechanisms will inactivate all but one X in each of her cells. As a result, X-chromosomal abnormalities are typically associated with mild mental and physical defects, as well as sterility. If the X chromosome is absent altogether, the individual will not develop.

Several errors in sex chromosome number have been characterized. Individuals with three X chromosomes, called triplo-X, appear female but express developmental delays and reduced fertility. The XXY chromosome complement, corresponding to one type of Klinefelter syndrome, corresponds to male individuals with small testes, enlarged breasts, and reduced body hair. The extra X chromosome undergoes inactivation to compensate for the excess genetic dosage. Turner syndrome, characterized as an X0 chromosome complement (i.e., only a single sex chromosome), corresponds to a female individual with short stature, webbed skin in the neck region, hearing and cardiac impairments, and sterility.

An individual with more than the correct number of chromosome sets (two for diploid species) is called polyploid. For instance, fertilization of an abnormal diploid egg with a normal haploid sperm would yield a triploid zygote. Polyploid animals are extremely rare, with only a few examples among the

flatworms, crustaceans, amphibians, fish, and lizards. Triploid animals are sterile because meiosis cannot proceed normally with an odd number of chromosome sets. In contrast, polyploidy is very common in the plant kingdom, and polyploid plants tend to be larger and more robust than euploids of their species.

Chromosome Structural Rearrangements

Cytologists have characterized numerous structural rearrangements in chromosomes, including partial duplications, deletions, inversions, and translocations. Duplications and deletions often produce offspring that survive but exhibit physical and mental abnormalities. Cri-du-chat (from the French for "cry of the cat") is a syndrome associated with nervous system abnormalities and identifiable physical features that results from a deletion of most of the small arm of chromosome 5 (Figure 7.11). Infants with this genotype emit a characteristic high-pitched cry upon which the disorder's name is based.

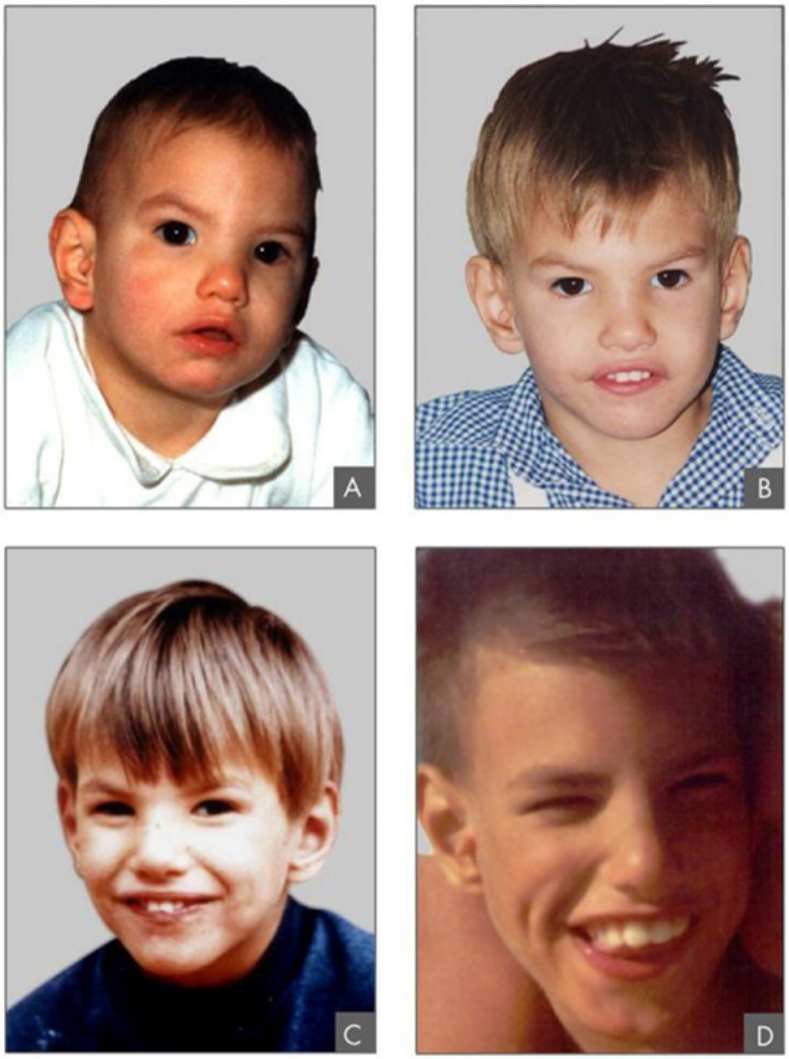


Figure 7.11 This individual with cri-du-chat syndrome is shown at various ages: (A) age two, (B) age four, (C) age nine, and (D) age 12. (credit: Paola Cerruti Mainardi)

Chromosome inversions and translocations can be identified by observing cells during meiosis because homologous chromosomes with a rearrangement in one of the pair must contort to maintain appropriate gene alignment and pair effectively during prophase I.

A chromosome inversion is the detachment, 180° rotation, and reinsertion of part of a chromosome. Unless they disrupt a gene sequence, inversions only change the orientation of genes and are likely to have more mild effects than aneuploid errors.

Evolution in Action

The Chromosome 18 InversionNot all structural rearrangements of chromosomes produce nonviable, impaired, or infertile individuals. In rare instances, such a change can result in the evolution of a new species. In fact, an inversion in chromosome 18 appears to have contributed to the evolution of humans. This inversion is not present in our closest genetic relatives, the chimpanzees.

The chromosome 18 inversion is believed to have occurred in early humans following their divergence from a common ancestor with chimpanzees approximately five million years ago. Researchers have suggested that a long stretch of DNA was duplicated on chromosome 18 of an ancestor to humans, but that during the duplication it was inverted (inserted into the chromosome in reverse orientation.

A comparison of human and chimpanzee genes in the region of this inversion indicates that two genes—*ROCK1* and *USP14*—are farther apart on human chromosome 18 than they are on the corresponding chimpanzee chromosome. This suggests that one of the inversion breakpoints occurred between these two genes. Interestingly, humans and chimpanzees express *USP14* at distinct levels in specific cell types, including cortical cells and fibroblasts. Perhaps the chromosome 18 inversion in an ancestral human repositioned specific genes and reset their expression levels in a useful way. Because both *ROCK1* and *USP14* code for enzymes, a change in their expression could alter cellular function. It is not known how this inversion contributed to hominid evolution, but it appears to be a significant factor in the divergence of humans from other primates.[1]

A translocation occurs when a segment of a chromosome dissociates and reattaches to a different, nonhomologous chromosome. Translocations can be benign or have devastating effects, depending on how the positions of genes are altered with respect to regulatory sequences. Notably, specific translocations have been associated with several cancers and with schizophrenia. Reciprocal translocations result from the exchange of chromosome segments between two nonhomologous chromosomes such that there is no gain or loss of genetic information (Figure 7.12).

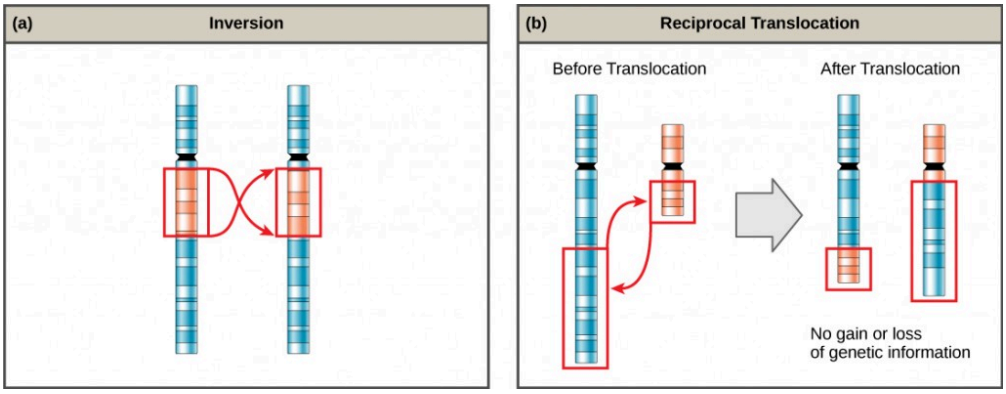


Figure 7.12 An (a) inversion occurs when a chromosome segment breaks from the chromosome, reverses its orientation, and then reattaches in the original position. A (b) reciprocal translocation occurs between two nonhomologous chromosomes and does not cause any genetic information to be lost or duplicated. (credit: modification of work by National Human Genome Research Institute (USA)

Section Summary

The number, size, shape, and banding pattern of chromosomes make them easily identifiable in a karyogram and allow for the assessment of many chromosomal abnormalities. Disorders in chromosome number, or aneuploidies, are typically lethal to the embryo, although a few trisomic genotypes are viable. Because of X inactivation, aberrations in sex chromosomes typically have milder effects on an individual. Aneuploidies also include instances in which segments of a chromosome are duplicated or deleted. Chromosome structures also may be rearranged, for example by inversion or translocation. Both of these aberrations can result in negative effects on development, or death. Because they force chromosomes to assume contorted pairings during meiosis I, inversions and translocations are often associated with reduced fertility because of the likelihood of nondisjunction.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4532#h5p-46

Glossary

aneuploid: an individual with an error in chromosome number; includes deletions and duplications of chromosome segments

autosome: any of the non-sex chromosomes

chromosome inversion: the detachment, 180° rotation, and reinsertion of a chromosome arm

euploid: an individual with the appropriate number of chromosomes for their species

karyogram: the photographic image of a karyotype

karyotype: the number and appearance of an individuals chromosomes, including the size, banding patterns, and centromere position

monosomy: an otherwise diploid genotype in which one chromosome is missing

nondisjunction: the failure of synapsed homologs to completely separate and migrate to separate poles during the first cell division of meiosis

polyploid: an individual with an incorrect number of chromosome sets

translocation: the process by which one segment of a chromosome dissociates and reattaches to a different, nonhomologous chromosome

trisomy: an otherwise diploid genotype in which one entire chromosome is duplicated

X inactivation: the condensation of X chromosomes into Barr bodies during embryonic development in females to compensate for the double genetic dose

Footnotes

1 V Goidts, et al., "Segmental duplication associated with the human-specific inversion of chromosome 18: a further example of the impact of segmental duplications on karyotype and genome evolution in primates," *Human Genetics*, 115 (2004):116–22.

Chapter 8: Introduction to Patterns of Inheritance

Figure 8.1 Experimenting with thousands of garden peas, Mendel uncovered the fundamentals of genetics. (credit: modification of work by Jerry Kirkhart)

Genetics is the study of heredity. Johann Gregor Mendel set the framework for genetics long before chromosomes or genes had been identified, at a time when meiosis was not well understood. Mendel selected a simple biological system and conducted methodical, quantitative analyses using **large sample sizes**. Because of Mendel's work, the fundamental principles of heredity were revealed. We now know that genes, carried on chromosomes, are the basic functional units of heredity with the ability to be replicated, expressed, or mutated. Today, the postulates put forth by Mendel form the basis of classical, or Mendelian, genetics. Not all genes are transmitted from parents to offspring according to Mendelian genetics, but Mendel's experiments serve as an excellent starting point for thinking about inheritance.

Search for Key Points in Chapter 8

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4535#h5p-47

8.1 Mendel's Experiments

Learning Objectives

By the end of this section, you will be able to:

- Explain the scientific reasons for the success of Mendel's experimental work
- Describe the expected outcomes of monohybrid crosses involving dominant and recessive alleles.

Figure 8.2 Johann Gregor Mendel set the framework for the study of genetics.

Watch the interactive video

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4541#h5p-48

Johann Gregor Mendel (1822–1884) 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*,[1] 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 8.3). Mendel collected the 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.

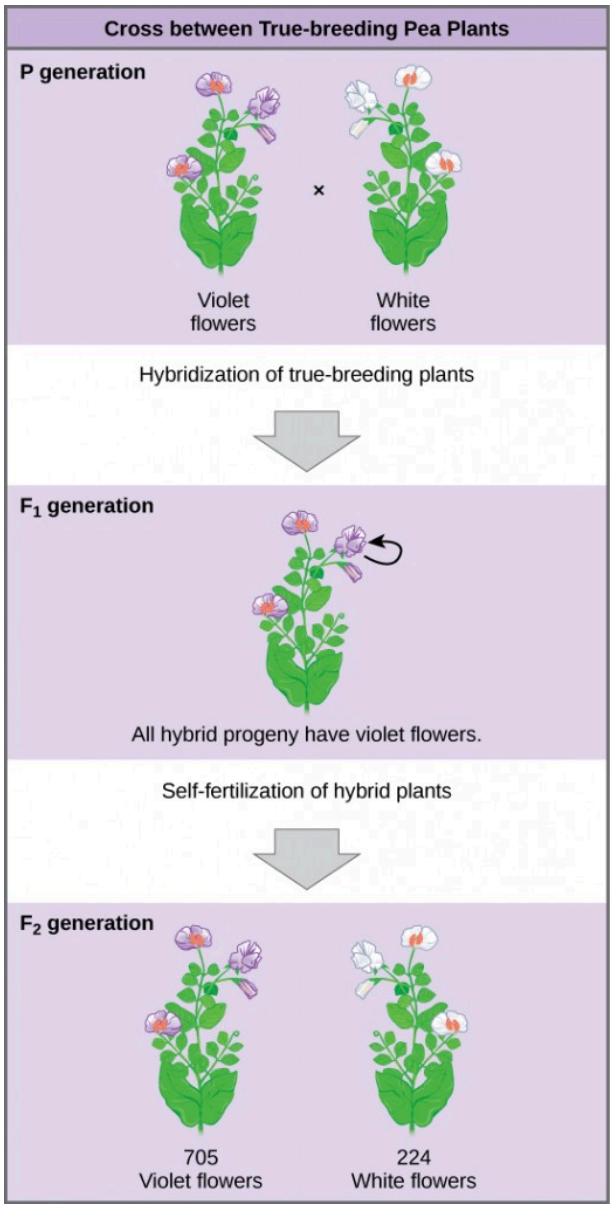


Figure 8.3 Mendel's process for performing crosses included examining flower color.

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 8.4).

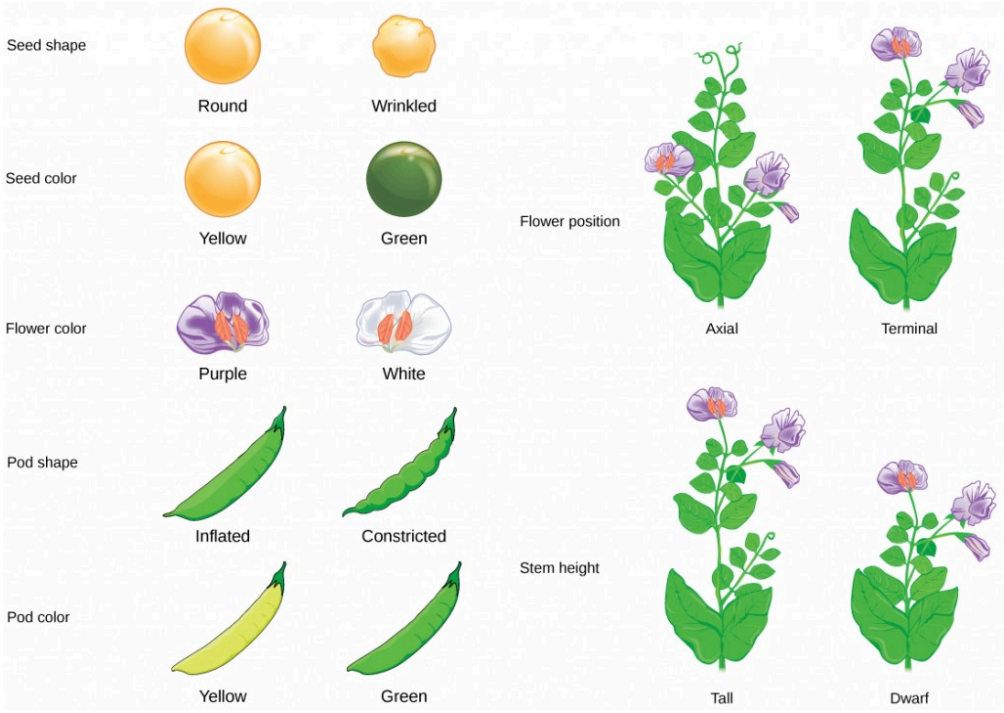


Figure 8.4 Mendel identified seven pea plant characteristics.

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.

Concept in Action

For an excellent review of Mendel's experiments and to perform your own crosses and identify patterns of inheritance, visit the Mendel's Peas web lab.

Section Summary

Working with garden pea plants, Mendel found that crosses between parents that differed for one trait produced F_1 offspring that all expressed one parent's traits. The traits that were visible in the F_1 generation are referred to as dominant, and traits that disappear in the F_1 generation are described as recessive. When the F_1 plants in Mendel's experiment were self-crossed, the F_2 offspring exhibited the dominant trait or the recessive trait in a 3:1 ratio, confirming that the recessive trait had been transmitted faithfully from the original P parent. Reciprocal crosses generated identical F_1 and F_2 offspring ratios. By examining sample sizes, Mendel showed that traits were inherited as independent events.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4541#h5p-49

Glossary

continuous variation: a variation in a characteristic in which individuals show a range of traits with small differences between them

discontinuous variation: a variation in a characteristic in which individuals show two, or a few, traits with large differences between them

dominant: describes a trait that masks the expression of another trait when both versions of the gene are present in an individual

F_1: the first filial generation in a cross; the offspring of the parental generation

F_2: the second filial generation produced when F_1 individuals are self-crossed or fertilized with each other

hybridization: the process of mating two individuals that differ, with the goal of achieving a certain characteristic in their offspring

model system: a species or biological system used to study a specific biological phenomenon to gain understanding that will be applied to other species

P: the parental generation in a cross

recessive: describes a trait whose expression is masked by another trait when the alleles for both traits are present in an individual

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

trait: a variation in an inherited characteristic

Footnotes

1 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]

8.2 Laws of Inheritance

Learning Objectives

By the end of this section, you will be able to:

- Explain the relationship between genotypes and phenotypes in dominant and recessive gene systems
- Use a Punnett square to calculate the expected proportions of genotypes and phenotypes in a monohybrid cross
- Explain Mendel's law of segregation and independent assortment in terms of genetics and the events of meiosis
- Explain the purpose and methods of a test cross

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 8.5). 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*.

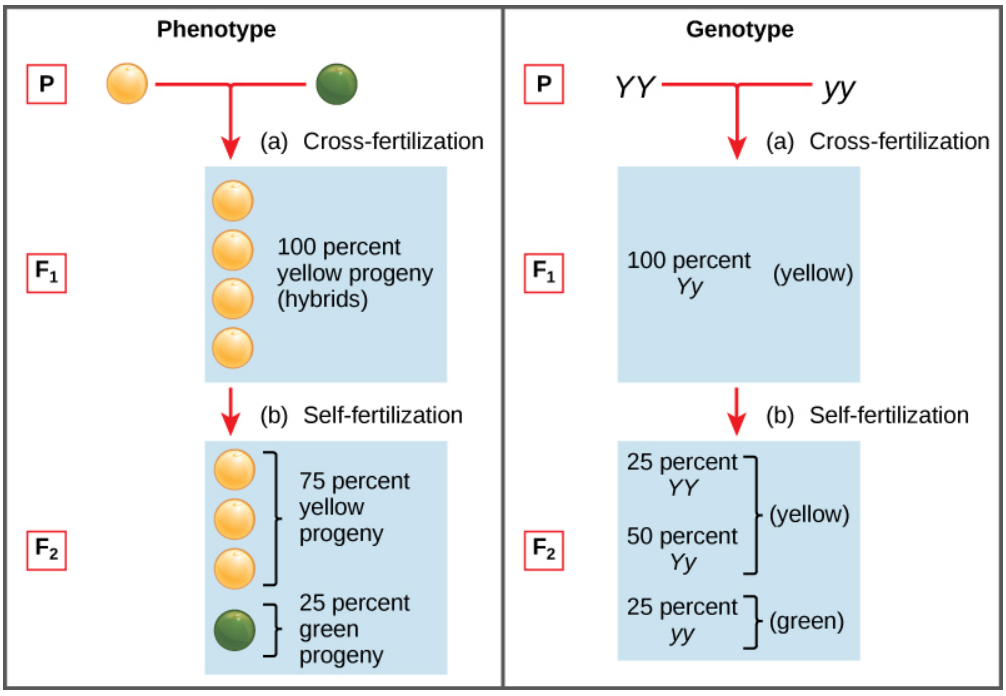


Figure 8.5 Phenotypes are physical expressions of traits that are transmitted by alleles. Capital letters represent dominant alleles and lowercase letters represent recessive alleles. The phenotypic ratios are the ratios of visible characteristics. The genotypic ratios are the ratios of gene combinations 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.

Correspondence between Genotype and Phenotype for a Dominant-Recessive Characteristic.

	Homozygous	**Heterozygous**	**Homozygous**
Genotype	*YY*	*Yy*	*yy*
Phenotype	yellow	yellow	green

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 8.6), and these offspring will breed true when self-crossed.

Figure 8.6 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 8.9 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 8.7).

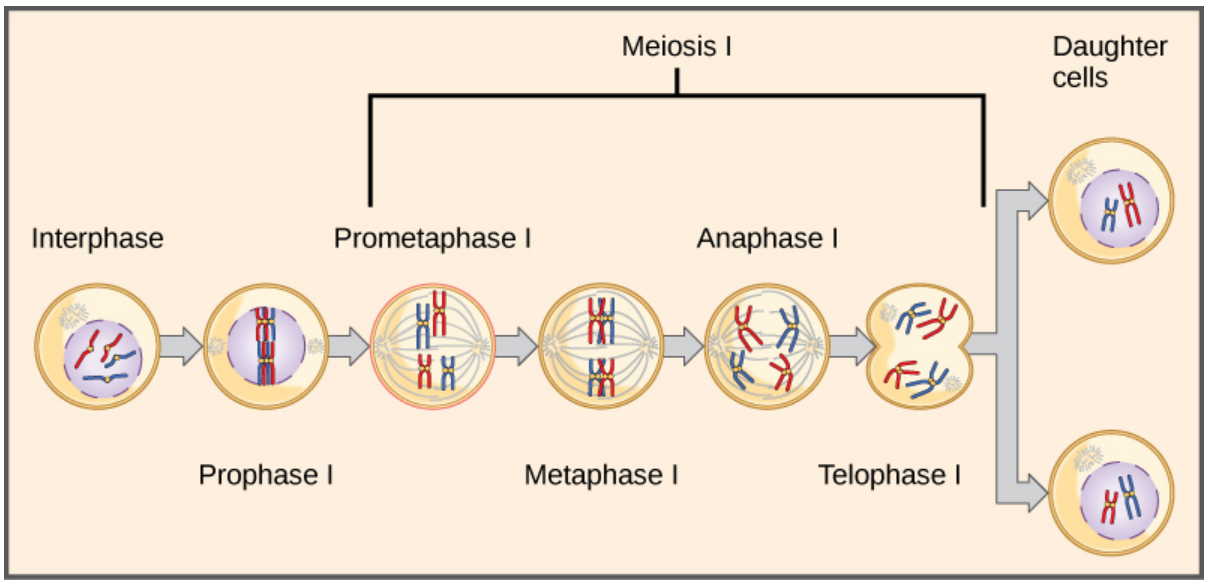


Figure 8.7 The first division in meiosis is shown.

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 8.8). 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 8.9). The test cross further validates Mendel's postulate that pairs of unit factors segregate equally.

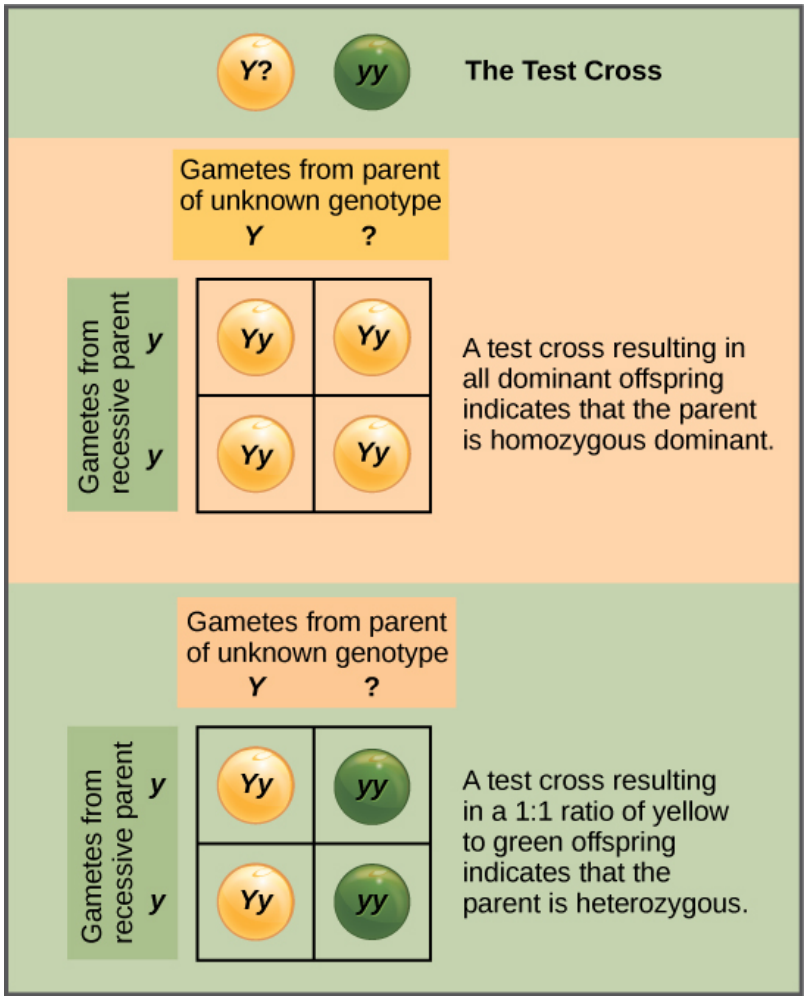


Figure 8.8 A test cross can be performed to determine whether an organism expressing a dominant trait is a homozygote or a heterozygote.

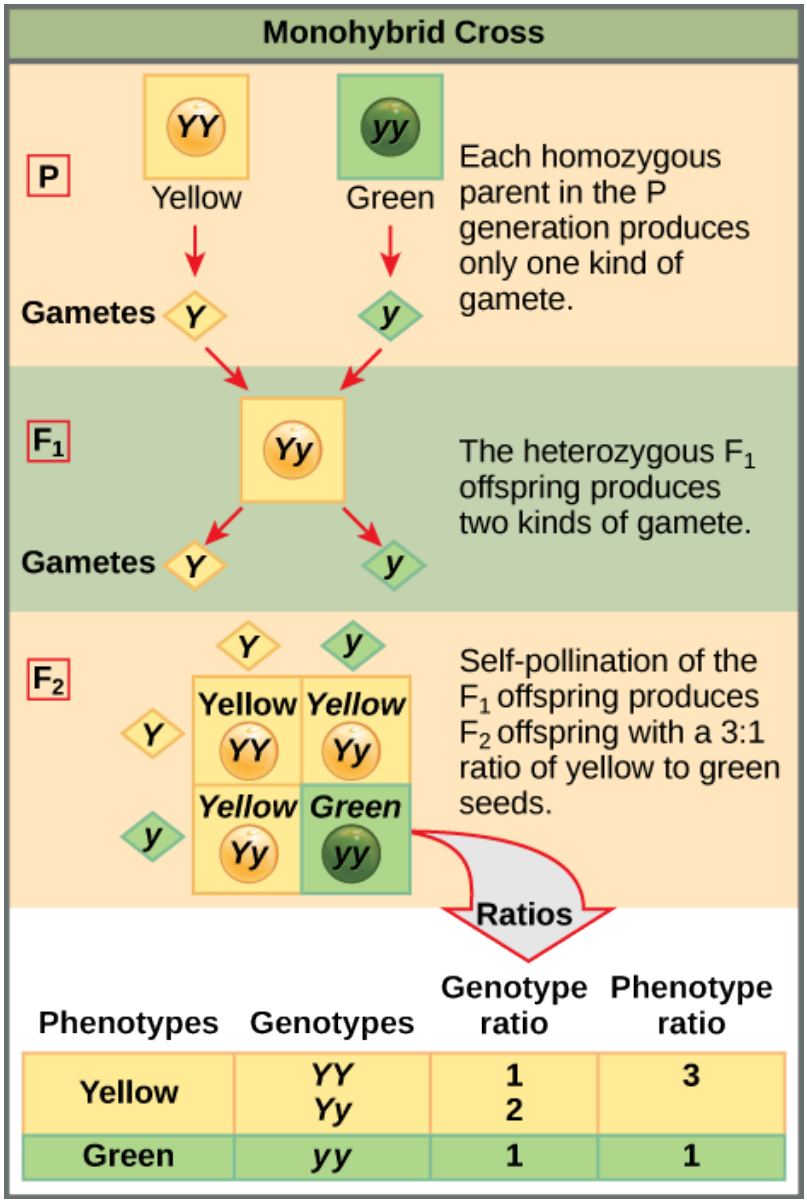


Figure 8.9 This Punnett square shows the cross between plants with yellow seeds and green seeds. The cross between the true-breeding P plants produces F1 heterozygotes that can be self-fertilized. The self-cross of the F1 generation can be analyzed with a Punnett square to predict the genotypes of the F2 generation. Given an inheritance pattern of dominant–recessive, the genotypic and phenotypic ratios can then be determined.

In pea plants, round peas (*R*) are dominant to wrinkled peas (*r*). You do a test cross between a pea plant with wrinkled peas (genotype *rr*) and a plant of unknown genotype that has round peas. You end up with three plants, all which have round peas. From this data, can you tell if the parent plant is homozygous dominant or heterozygous?

You cannot be sure if the plant is homozygous or heterozygous as the data set is too small: by random

chance, all three plants might have acquired only the dominant gene even if the recessive one is present.

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 8.10).

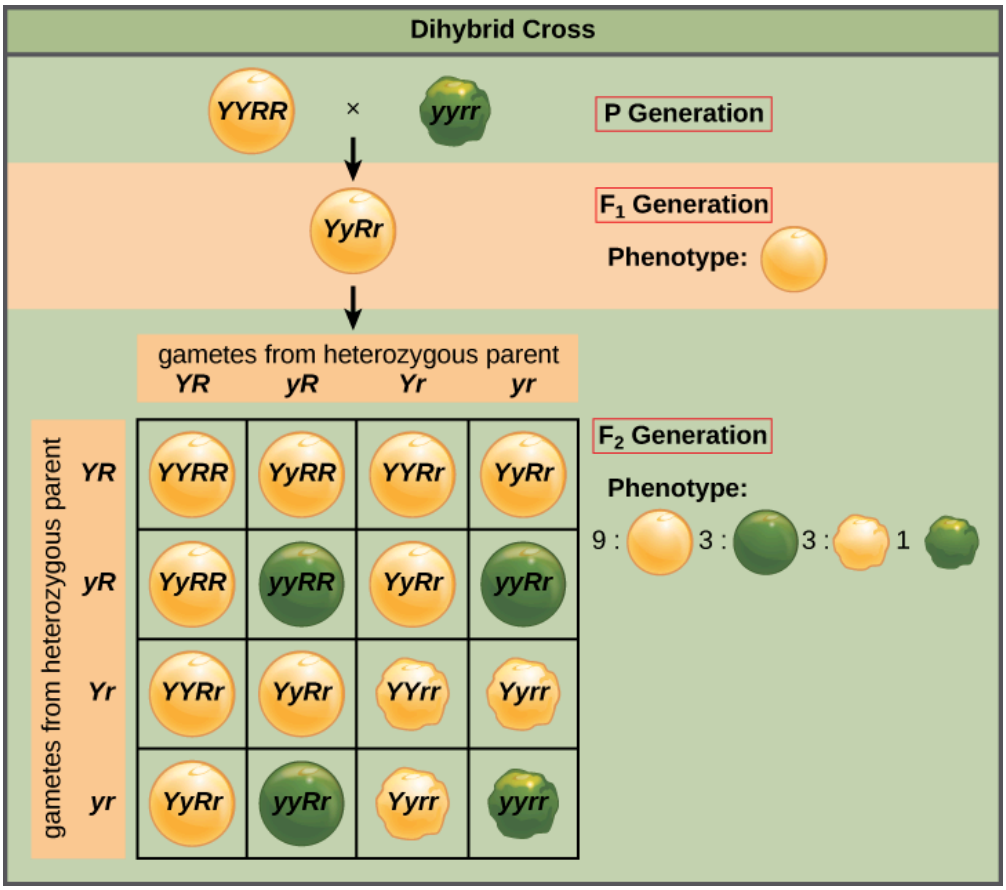


Figure 8.10 A dihybrid cross in pea plants involves the genes for seed color and texture. The P cross produces F1 offspring that are all heterozygous for both characteristics. The resulting 9:3:3:1 F2 phenotypic ratio is obtained using a Punnett square.

In pea plants, purple flowers (*P*) are dominant to white (*p*), and yellow peas (*Y*) are dominant to green (*y*). What are the possible genotypes and phenotypes for a cross between *PpYY* and *ppYy* pea plants? How many squares would you need to complete a Punnett square analysis of this cross?

The possible genotypes are PpYY, PpYy, ppYY, and ppYy. The former two genotypes would result in

plants with purple flowers and yellow peas, while the latter two genotypes would result in plants with white flowers with yellow peas, for a 1:1 ratio of each phenotype. You only need a 2 × 2 Punnett square (four squares total) to do this analysis because two of the alleles are homozygous.

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 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. 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 8.11).

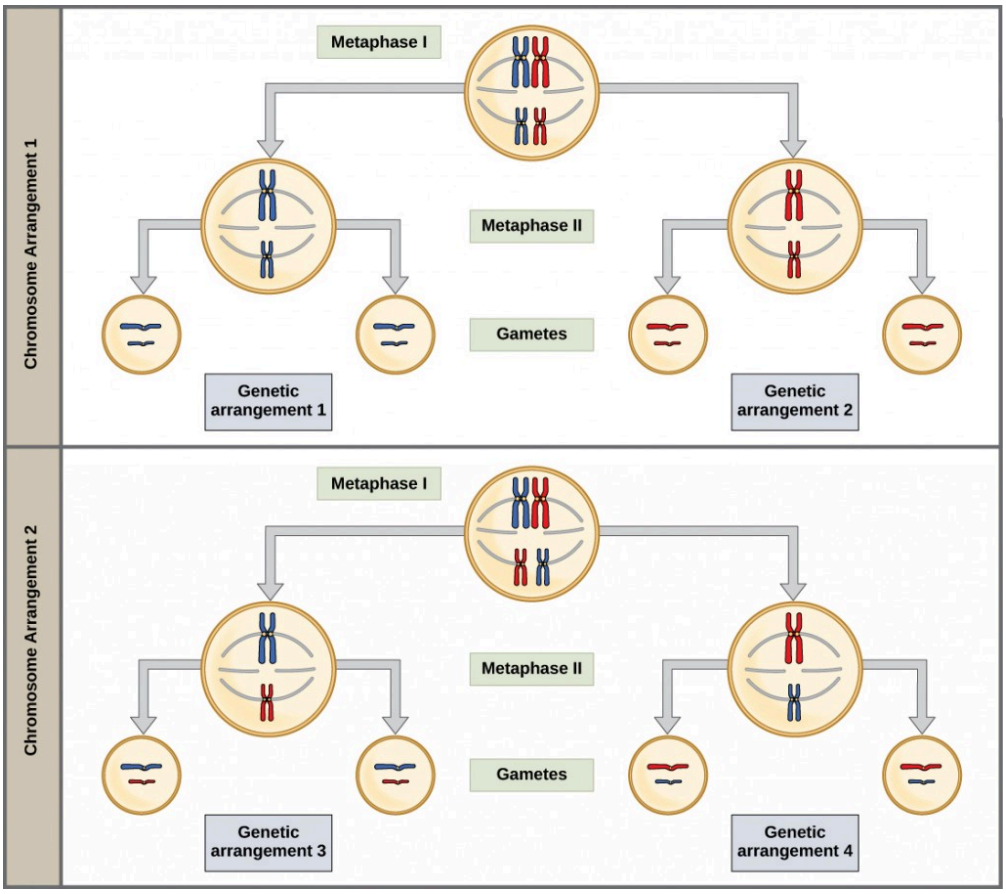


Figure 8.11 The random segregation into daughter nuclei that happens during the first division in meiosis can lead to a variety of possible genetic arrangements.

Probability Basics

Probabilities are mathematical measures of likelihood. The empirical probability of an event is calculated by dividing the number of times the event occurs by the total number of opportunities for the event to occur. It is also possible to calculate theoretical probabilities by dividing the number of times that an event is expected to occur by the number of times that it could occur. Empirical probabilities come from observations, like those of Mendel. Theoretical probabilities come from knowing how the events are produced and assuming that the probabilities of individual outcomes are equal. A probability of one for some event indicates that it is guaranteed to occur, whereas a probability of zero indicates that it is guaranteed not to occur. An example of a genetic event is a round seed produced by a pea plant. In his experiment, Mendel demonstrated that the probability of the event "round seed" occurring was one in the F_1 offspring of true-breeding parents, one of which has round seeds and one of which has wrinkled seeds. When the F_1 plants were subsequently self-crossed, the probability of any given F_2 offspring having round seeds was now three out of four. In other words, in a large population of F_2 offspring chosen at random, 75 percent were expected to have round seeds, whereas 25 percent were expected to have wrinkled seeds. Using large numbers of crosses, Mendel was able to calculate probabilities and use these to predict the outcomes of other crosses.

The Product Rule and Sum Rule

Mendel demonstrated that the pea-plant characteristics he studied were transmitted as discrete units from parent to offspring. As will be discussed, Mendel also determined that different characteristics, like seed color and seed texture, were transmitted independently of one another and could be considered in separate probability analyses. For instance, performing a cross between a plant with green, wrinkled seeds and a plant with yellow, round seeds still produced offspring that had a 3:1 ratio of green:yellow seeds (ignoring seed texture) and a 3:1 ratio of round:wrinkled seeds (ignoring seed color). The characteristics of color and texture did not influence each other.

The product rule of probability can be applied to this phenomenon of the independent transmission of characteristics. The product rule states that the probability of two independent events occurring together can be calculated by multiplying the individual probabilities of each event occurring alone. To demonstrate the product rule, imagine that you are rolling a six-sided die (D) and flipping a penny (P) at the same time. The die may roll any number from 1–6 ($D_{\#}$), whereas the penny may turn up heads (P_H) or tails (P_T). The outcome of rolling the die has no effect on the outcome of flipping the penny and vice versa. There are 12 possible outcomes of this action, and each event is expected to occur with equal probability.

Twelve Equally Likely Outcomes of Rolling a Die and Flipping a Penny

Rolling Die	Flipping Penny
D_1	P_H
D_1	P_T
D_2	P_H
D_2	P_T
D_3	P_H
D_3	P_T
D_4	P_H
D_4	P_T
D_5	P_H
D_5	P_T
D_6	P_H
D_6	P_T

Of the 12 possible outcomes, the die has a 2/12 (or 1/6) probability of rolling a two, and the penny has a 6/12 (or 1/2) probability of coming up heads. By the product rule, the probability that you will obtain the combined outcome 2 and heads is: (D_2) x (P_H) = (1/6) x (1/2) or 1/12. Notice the word "and" in the description of the probability. The "and" is a signal to apply the product rule. For example, consider how the product rule is applied to the dihybrid cross: the probability of having both dominant traits in the F_2 progeny is the product of the probabilities of having the dominant trait for each characteristic, as shown here:

$$3/4 \times 3/4 = 9/16$$

On the other hand, the sum rule of probability is applied when considering two mutually exclusive outcomes that can come about by more than one pathway. The sum rule states that the probability of the occurrence of one event or the other event, of two mutually exclusive events, is the sum of their individual probabilities. Notice the word "or" in the description of the probability. The "or" indicates that you should apply the sum rule. In this case, let's imagine you are flipping a penny (P) and a quarter (Q). What is the probability of one coin coming up heads and one coin coming up tails? This outcome can be achieved by two cases: the penny may be heads (P_H) and the quarter may be tails (Q_T), or the quarter may be heads (Q_H) and the penny may be tails (P_T). Either case fulfills the outcome. By the sum rule, we calculate the probability of obtaining one head and one tail as $[(P_H) \times (Q_T)] + [(Q_H) \times (P_T)] = [(1/2) \times (1/2)] + [(1/2) \times (1/2)] = 1/2$. You should also notice that we used the product rule to calculate the probability of P_H and Q_T, and also the probability of P_T and Q_H, before we summed them. Again, the sum rule can be applied to show the probability of having just one dominant trait in the F_2 generation of a dihybrid cross:

$$3/16 + 3/4 = 15/16$$

The Product Rule and Sum Rule

Product Rule	Sum Rule
For independent events A and B, the probability (P) of them both occurring (A *and* B) is ($P_A \times P_B$)	For mutually exclusive events A and B, the probability (P) that at least one occurs (A *or* B) is ($P_A + P_B$)

To use probability laws in practice, it is necessary to work with large sample sizes because small sample sizes are prone to deviations caused by chance. The large quantities of pea plants that Mendel examined allowed him calculate the probabilities of the traits appearing in his F_2 generation. As you will learn, this discovery meant that when parental traits were known, the offspring's traits could be predicted accurately even before fertilization.

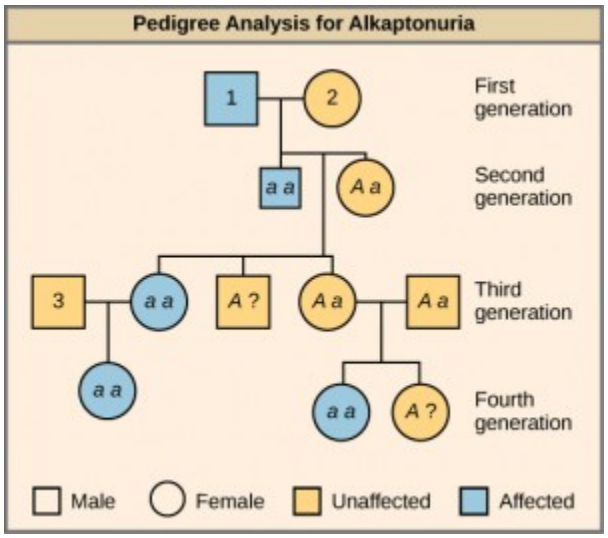


Figure 8.12

Alkaptonuria is a recessive genetic disorder in which two amino acids, phenylalanine and tyrosine, are not properly metabolized. Affected individuals may have darkened skin and brown urine, and may suffer joint damage and other complications. In this pedigree, individuals with the disorder are indicated in blue and have the genotype *aa*. Unaffected individuals are indicated in yellow and have the genotype *AA* or *Aa*. Note that it is often possible to determine a person's genotype from the genotype of their offspring. For example, if neither parent has the disorder but their child does, they must be heterozygous. Two individuals on the pedigree have an unaffected phenotype but unknown genotype. Because they do not have the disorder, they must have at least one normal allele, so their genotype gets the "*A?*" designation.

What are the genotypes of the individuals labeled 1, 2 and 3?

Section Summary

When true-breeding, or homozygous, individuals that differ for a certain trait are crossed, all of the offspring will be heterozygous for that trait. If the traits are inherited as dominant and recessive, the F_1

offspring will all exhibit the same phenotype as the parent homozygous for the dominant trait. If these heterozygous offspring are self-crossed, the resulting F_2 offspring will be equally likely to inherit gametes carrying the dominant or recessive trait, giving rise to offspring of which one quarter are homozygous dominant, half are heterozygous, and one quarter are homozygous recessive. Because homozygous dominant and heterozygous individuals are phenotypically identical, the observed traits in the F_2 offspring will exhibit a ratio of three dominant to one recessive.

Mendel postulated that genes (characteristics) are inherited as pairs of alleles (traits) that behave in a dominant and recessive pattern. Alleles segregate into gametes such that each gamete is equally likely to receive either one of the two alleles present in a diploid individual. In addition, genes are assorted into gametes independently of one another. That is, in general, alleles are not more likely to segregate into a gamete with a particular allele of another gene.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4551#h5p-50

Glossary

allele: one of two or more variants of a gene that determines a particular trait for a characteristic

dihybrid: the result of a cross between two true-breeding parents that express different traits for two characteristics

genotype: the underlying genetic makeup, consisting of both physically visible and non-expressed alleles, of an organism

heterozygous: having two different alleles for a given gene on the homologous chromosomes

homozygous: having two identical alleles for a given gene on the homologous chromosomes

law of dominance: in a heterozygote, one trait will conceal the presence of another trait for the same characteristic

law of independent assortment: genes do not influence each other with regard to sorting of alleles into gametes; every possible combination of alleles is equally likely to occur

law of segregation: paired unit factors (i.e., genes) segregate equally into gametes such that offspring have an equal likelihood of inheriting any combination of factors

monohybrid: the result of a cross between two true-breeding parents that express different traits for only one characteristic

phenotype: the observable traits expressed by an organism

Punnett square: a visual representation of a cross between two individuals in which the gametes of each individual are denoted along the top and side of a grid, respectively, and the possible zygotic genotypes are recombined at each box in the grid

test cross: a cross between a dominant expressing individual with an unknown genotype and a homozygous recessive individual; the offspring phenotypes indicate whether the unknown parent is heterozygous or homozygous for the dominant trait

8.3 Extensions of the Laws of Inheritance

Learning Objectives

By the end of this section, you will be able to:

- Identify non-Mendelian inheritance patterns such as incomplete dominance, codominance, multiple alleles, and sex linkage from the results of crosses
- Explain the effect of linkage and recombination on gamete genotypes
- Explain the phenotypic outcomes of epistatic effects among genes
- Explain polygenic inheritance

Mendel studied traits with only one mode of inheritance 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. There are several important modes of inheritance, discovered after Mendel's work, that do not follow the dominant and recessive, single-gene model.

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 8.13), 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 8.13 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 8.14).

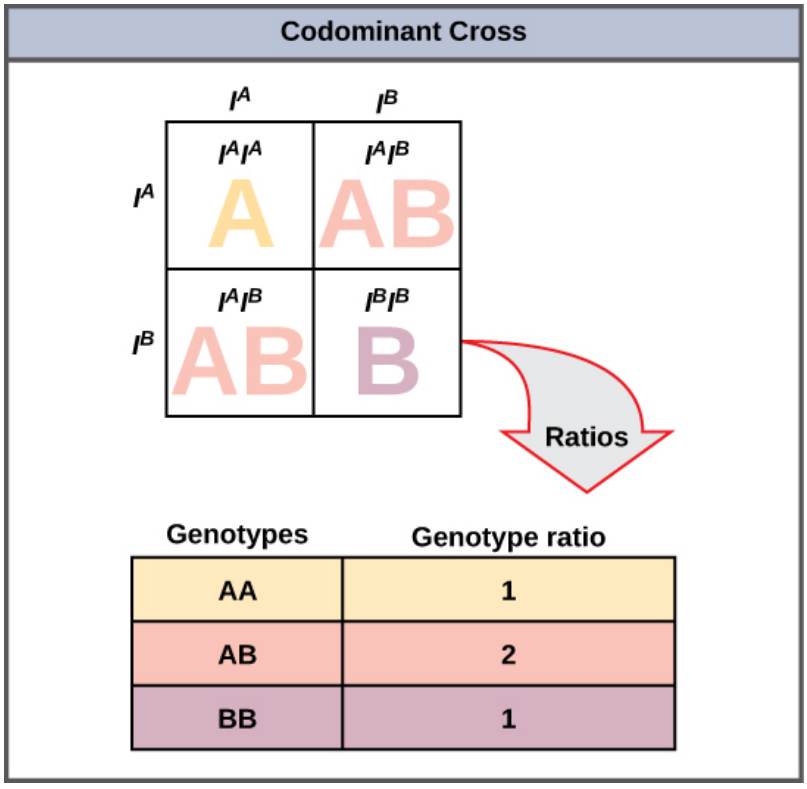


Figure 8.14 This Punnett square shows an AB/AB blood type cross

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 8.15. 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.

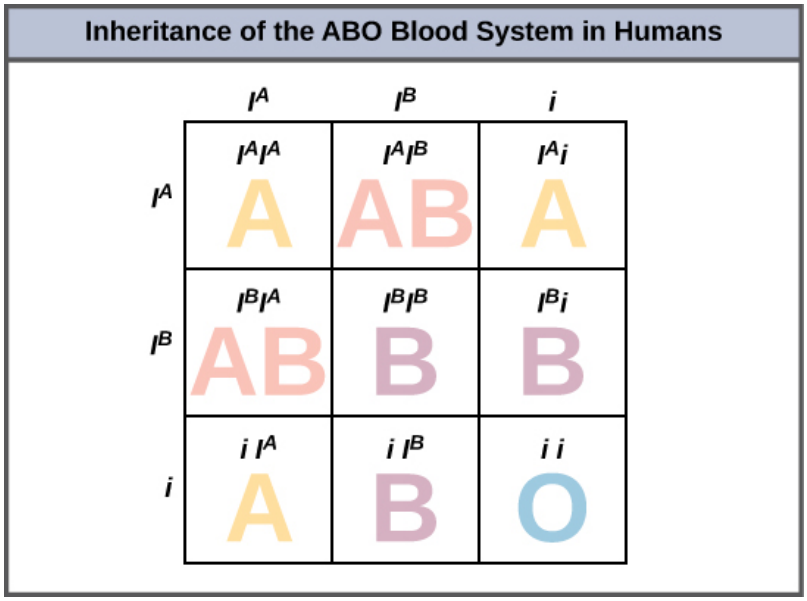


Figure 8.15 Inheritance of the ABO blood system in humans is shown.

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. In flies the wild-type eye color is red (X^W) and is dominant to white eye color (X^w) (Figure 8.16). 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 8.16 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 8.17). 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).

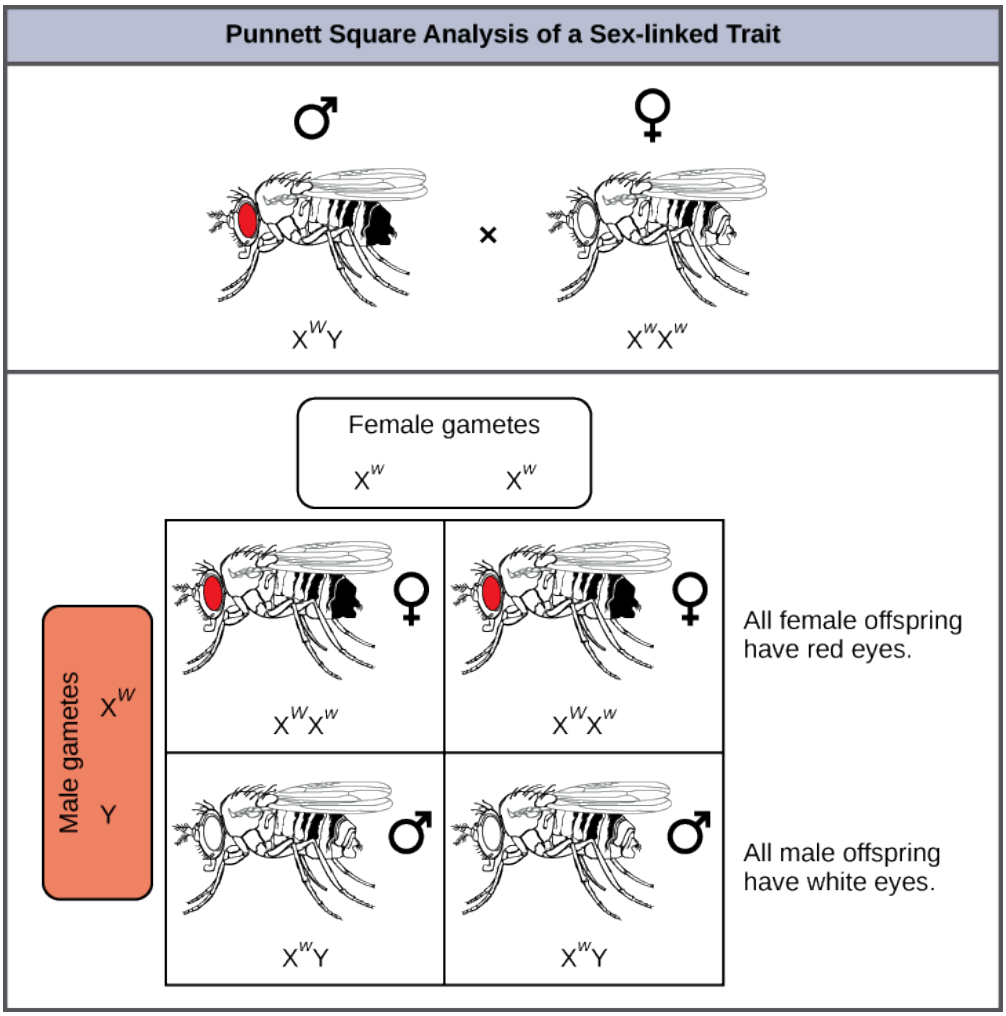


Figure 8.17 Crosses involving sex-linked traits often give rise to different phenotypes for the different sexes of offspring, as is the case for this cross involving red and white eye color in Drosophila. In the diagram, w is the white-eye mutant allele and W is the wild-type, red-eye allele.

What ratio of offspring would result from a cross between a white-eyed male and a female that is heterozygous for red eye color?

<!–

Half of the female offspring would be heterozygous (X^WX^w) with red eyes, and half would be homozygous recessive (X^wX^w) with white eyes. Half of the male offspring would be hemizygous

dominant (X^WY) with red eyes, and half would be hemizygous recessive (X^wY) with white eyes.
–>

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.

Concept in Action

Watch this video to learn more about sex-linked traits.

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 8.18). 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.

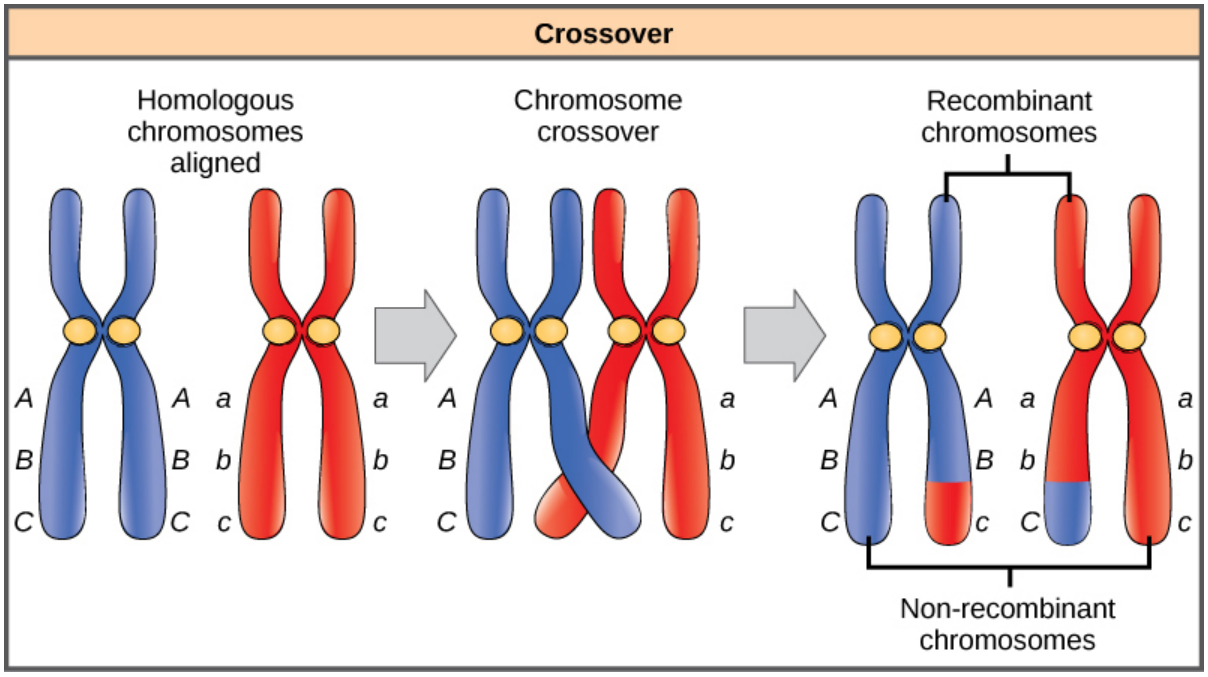


Figure 8.18 The process of crossover, or recombination, occurs when two homologous chromosomes align and exchange a segment of genetic material.

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.

Epistasis

Mendel's studies in pea plants implied that the sum of an individual's phenotype was controlled by genes (or as he called them, unit factors), such that every characteristic was distinctly and completely controlled by a single gene. In fact, single observable characteristics are almost always under the influence of multiple genes (each with two or more alleles) acting in unison. For example, at least eight genes contribute to eye color in humans.

Concept in Action

Eye color in humans is determined by multiple alleles. Use the Eye Color Calculator to predict the eye color of children from parental eye color.

In some cases, several genes can contribute to aspects of a common phenotype without their gene products ever directly interacting. In the case of organ development, for instance, genes may be expressed sequentially, with each gene adding to the complexity and specificity of the organ. Genes may function in complementary or synergistic fashions, such that two or more genes expressed simultaneously affect a phenotype. An apparent example of this occurs with human skin color, which appears to involve the action of at least three (and probably more) genes. Cases in which inheritance for a characteristic like skin color or human height depend on the combined effects of numerous genes are called polygenic inheritance.

Genes may also oppose each other, with one gene suppressing the expression of another. In epistasis, the interaction between genes is antagonistic, such that one gene masks or interferes with the expression of another. "Epistasis" is a word composed of Greek roots meaning **"standing upon."** The alleles that are being **masked or silenced** are said to be hypostatic to the epistatic alleles that are doing the masking. Often the biochemical basis of epistasis is a gene pathway in which expression of one gene is dependent on the function of a gene that precedes or follows it in the pathway.

An example of epistasis is pigmentation in mice. The wild-type coat color, agouti (AA) is dominant to solid-colored fur (aa). However, a separate gene C, when present as the recessive homozygote (cc), negates any expression of pigment from the A gene and results in an albino mouse (Figure 8.19). Therefore, the genotypes *AAcc*, *Aacc*, and *aacc* all produce the same albino phenotype. A cross

between heterozygotes for both genes (*AaCc* x *AaCc*) would generate offspring with a phenotypic ratio of 9 agouti:3 black:4 albino (Figure 8.19). In this case, the *C* gene is epistatic to the *A* gene.

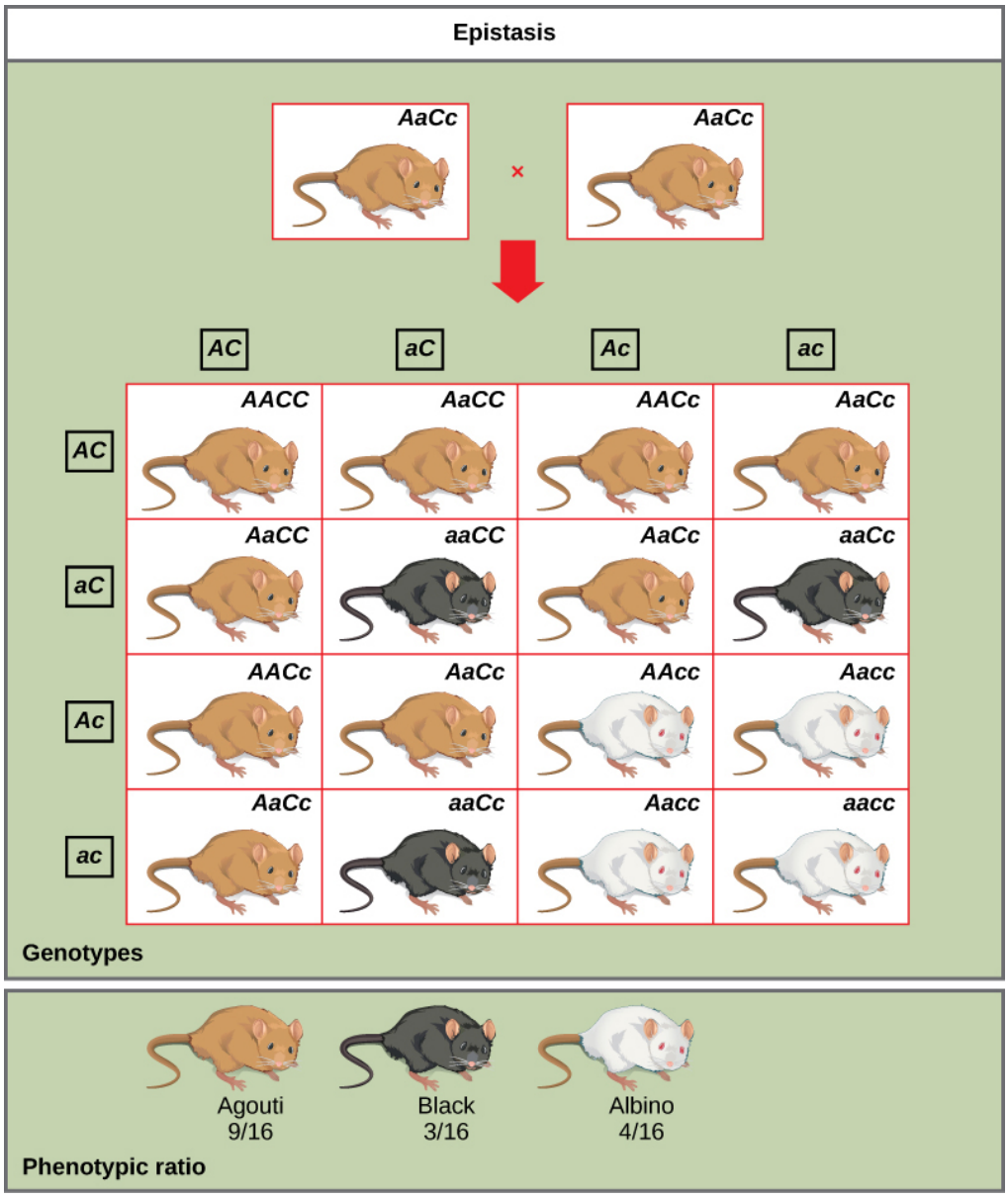


Figure 8.19 In this example of epistasis, one gene (C) masks the expression of another (A) for coat color. When the C allele is present, coat color is expressed; when it is absent (cc), no coat color is expressed. Coat color depends on the A gene, which shows dominance, with the recessive homozygote showing a different phenotype than the heterozygote or dominant homozygote.

Section Summary

Alleles do not always behave in dominant and recessive patterns. Incomplete dominance describes situations in which the heterozygote exhibits a phenotype that is intermediate between the homozygous phenotypes. Codominance describes the simultaneous expression of both of the alleles in the heterozygote. Although diploid organisms can only have two alleles for any given gene, it is common

for more than two alleles for a gene to exist in a population. In humans, as in many animals and some plants, females have two X chromosomes and males have one X and one Y chromosome. Genes that are present on the X but not the Y chromosome are said to be X-linked, such that males only inherit one allele for the gene, and females inherit two.

According to Mendel's law of independent assortment, genes sort independently of each other into gametes during meiosis. This occurs because chromosomes, on which the genes reside, assort independently during meiosis and crossovers cause most genes on the same chromosomes to also behave independently. When genes are located in close proximity on the same chromosome, their alleles tend to be inherited together. This results in offspring ratios that violate Mendel's law of independent assortment. However, recombination serves to exchange genetic material on homologous chromosomes such that maternal and paternal alleles may be recombined on the same chromosome. This is why alleles on a given chromosome are not always inherited together. Recombination is a random event occurring anywhere on a chromosome. Therefore, genes that are far apart on the same chromosome are likely to still assort independently because of recombination events that occurred in the intervening chromosomal space.

Whether or not they are sorting independently, genes may interact at the level of gene products, such that the expression of an allele for one gene masks or modifies the expression of an allele for a different gene. This is called epistasis.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4562#h5p-51

Glossary

codominance: in a heterozygote, complete and simultaneous expression of both alleles for the same characteristic

epistasis: an interaction between genes such that one gene masks or interferes with the expression of another

hemizygous: the presence of only one allele for a characteristic, as in X-linkage; hemizygosity makes descriptions of dominance and recessiveness irrelevant

incomplete dominance: in a heterozygote, expression of two contrasting alleles such that the individual displays an intermediate phenotype

linkage: a phenomenon in which alleles that are located in close proximity to each other on the same chromosome are more likely to be inherited together

recombination: the process during meiosis in which homologous chromosomes exchange linear segments of genetic material, thereby dramatically increasing genetic variation in the offspring and separating linked genes

wild type: the most commonly occurring genotype or phenotype for a given characteristic found in a population

X-linked: a gene present on the X chromosome, but not the Y chromosome

Footnotes

1 Sumiti Vinayak et al., "Origin and Evolution of Sulfadoxine Resistant *Plasmodium falciparum*," *PLoS Pathogens* 6 (2010): e1000830.

UNIT 3: MOLECULAR BIOLOGY AND BIOTECHNOLOGY

Unit 3: Molecular Biology and Biotechnology includes the following chapters:

Chapter 9: Introduction to Molecular Biology

Figure 9.1 Dolly the sheep was the first cloned mammal. Photo shows Dolly the sheep, which has been stuffed and placed in a glass case.

The three letters “DNA” have now become associated with crime solving, paternity testing, human identification, and genetic testing. DNA can be retrieved from hair, blood, or saliva. With the exception of identical twins, each person’s DNA is unique and it is possible to detect differences between human beings on the basis of their unique DNA sequence.

DNA analysis has many practical applications beyond forensics and paternity testing. DNA testing is used for tracing genealogy and identifying pathogens. In the medical field, DNA is used in diagnostics, new vaccine development, and cancer therapy. It is now possible to determine predisposition to many diseases by analyzing genes.

DNA is the genetic material passed from parent to offspring for all life on Earth. The technology of molecular genetics developed in the last half century has enabled us to see deep into the history of life to deduce the relationships between living things in ways never thought possible. It also allows us to understand the workings of evolution in populations of organisms. Over a thousand species have had their entire genome sequenced, and there have been thousands of individual human genome sequences completed. These sequences will allow us to understand human disease and the relationship of humans to the rest of the tree of life. Finally, molecular genetics techniques have revolutionized plant and animal breeding for human agricultural needs. All of these advances in biotechnology depended on basic research leading to the discovery of the structure of DNA in 1953, and the research since then that has uncovered the details of DNA replication and the complex process leading to the expression of DNA in the form of proteins in the cell.

Search for Key Points in Chapter 9

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4567#h5p-52

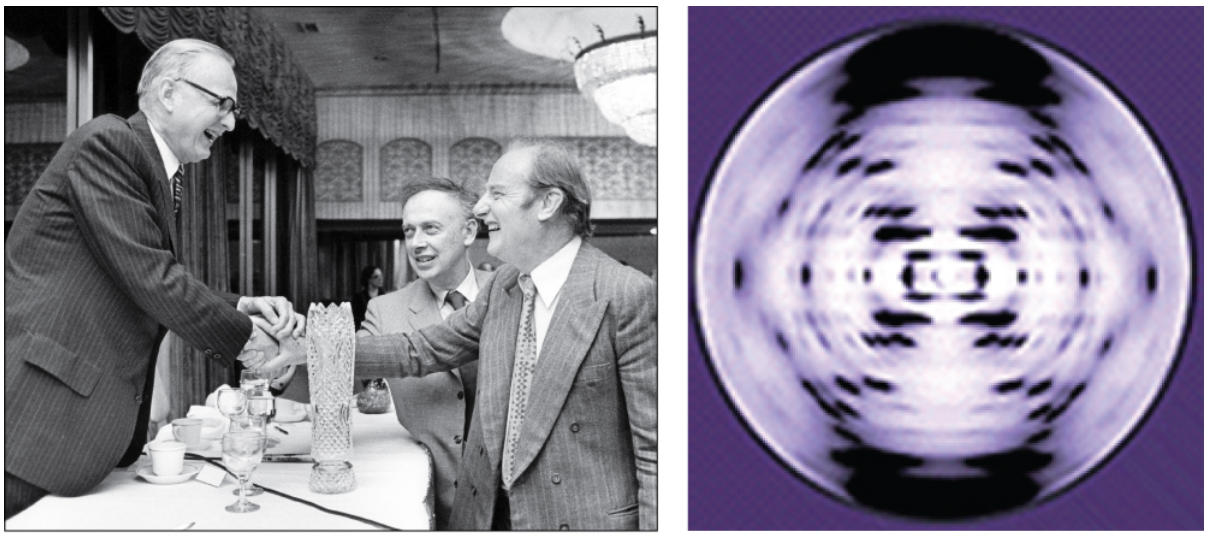
9.1 The Structure of DNA

Learning Objectives

By the end of this section, you will be able to:

- Describe the structure of DNA
- Describe how eukaryotic and prokaryotic DNA is arranged in the cell

In the 1950s, Francis Crick and James Watson worked together at the University of Cambridge, England, to determine the structure of DNA. Other scientists, such as Linus Pauling and Maurice Wilkins, were also actively exploring this field. Pauling had discovered the secondary structure of proteins using X-ray crystallography. X-ray crystallography is a method for investigating molecular structure by observing the patterns formed by X-rays shot through a crystal of the substance. The patterns give important information about the structure of the molecule of interest. In Wilkins' lab, researcher Rosalind Franklin was using X-ray crystallography to understand the structure of DNA. Watson and Crick were able to piece together the puzzle of the DNA molecule using Franklin's data (Figure 9.2). Watson and Crick also had key pieces of information available from other researchers such as Chargaff's rules. Chargaff had shown that of the four kinds of monomers (nucleotides) present in a DNA molecule, two types were always present in equal amounts and the remaining two types were also always present in equal amounts. This meant they were always paired in some way. In 1962, James Watson, Francis Crick, and Maurice Wilkins were awarded the Nobel Prize in Medicine for their work in determining the structure of DNA.



(a) (b)

Figure 9.2 Pioneering scientists (a) James Watson and Francis Crick are pictured here with American geneticist Maclyn McCarty. Scientist Rosalind Franklin discovered (b) the X-ray diffraction pattern of DNA, which helped to elucidate its double helix structure. (credit a: modification of work by Marjorie McCarty; b: modification of work by NIH)

Now let's consider the structure of the two types of nucleic acids, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). The building blocks of DNA are nucleotides, which are made up of three parts: a deoxyribose (5-carbon sugar), a phosphate group, and a nitrogenous base (Figure 9.3). There are four types of nitrogenous bases in DNA. Adenine (A) and guanine (G) are double-ringed purines, and cytosine (C) and thymine (T) are smaller, single-ringed pyrimidines. The nucleotide is named according to the nitrogenous base it contains.

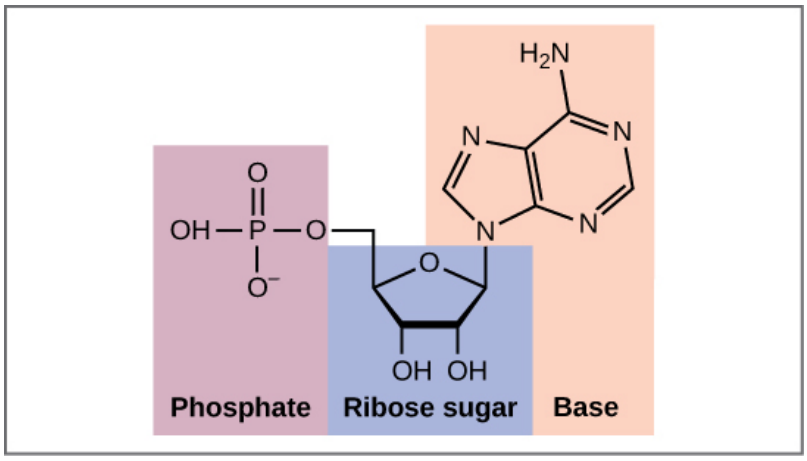


Figure 9.3 (a) Each DNA nucleotide is made up of a sugar, a phosphate group, and a base.

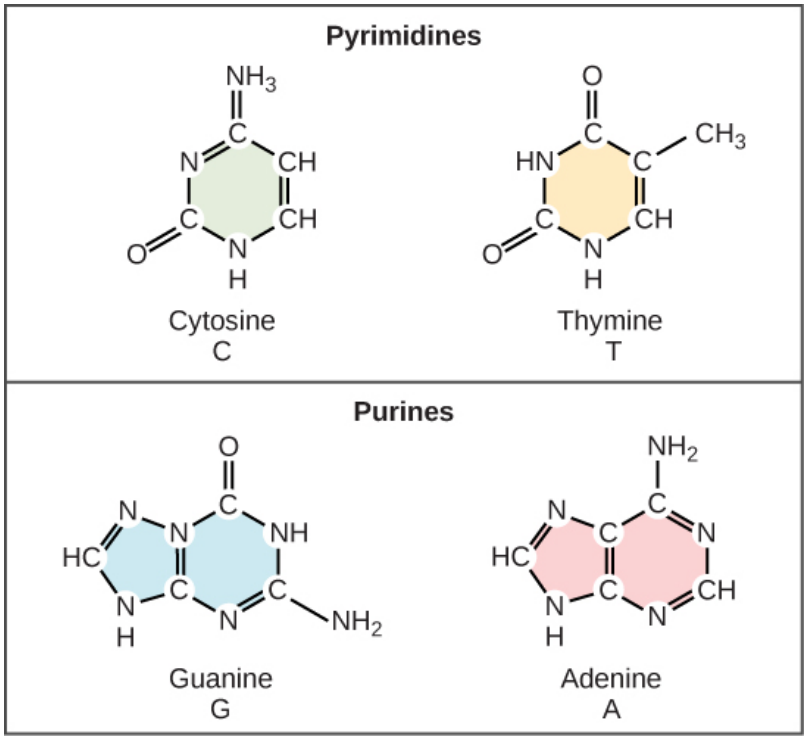


Figure 9.3 (b) Cytosine and thymine are pyrimidines. Guanine and adenine are purines.

The phosphate group of one nucleotide bonds covalently with the sugar molecule of the next nucleotide, and so on, forming a long polymer of nucleotide monomers. The sugar–phosphate groups line up in a "backbone" for each single strand of DNA, and the nucleotide bases stick out from this backbone. The carbon atoms of the five-carbon sugar are numbered clockwise from the oxygen as 1′, 2′, 3′, 4′, and 5′ (1′ is read as "one prime"). The phosphate group is attached to the 5′ carbon of one nucleotide and the 3′ carbon of the next nucleotide. In its natural state, each DNA molecule is actually composed of two single strands held together along their length with hydrogen bonds between the bases.

Watson and Crick proposed that the DNA is made up of two strands that are twisted around each other to form a right-handed helix, called a double helix. Base-pairing takes place between a purine and pyrimidine: namely, **A pairs with T, and G pairs with C**. In other words, adenine and thymine are complementary base pairs, and cytosine and guanine are also complementary base pairs. This is the basis for Chargaff's rule; because of their complementarity, there is as much adenine as thymine in a DNA molecule and as much guanine as cytosine. Adenine and thymine are connected by two hydrogen bonds, and cytosine and guanine are connected by three hydrogen bonds. The two strands are anti-parallel in nature; that is, one strand will have the 3′ carbon of the sugar in the "upward" position, whereas the other strand will have the 5′ carbon in the upward position. The diameter of the DNA double helix is uniform throughout because a purine (two rings) always pairs with a pyrimidine (one ring) and their combined lengths are always equal. (Figure 9.4).

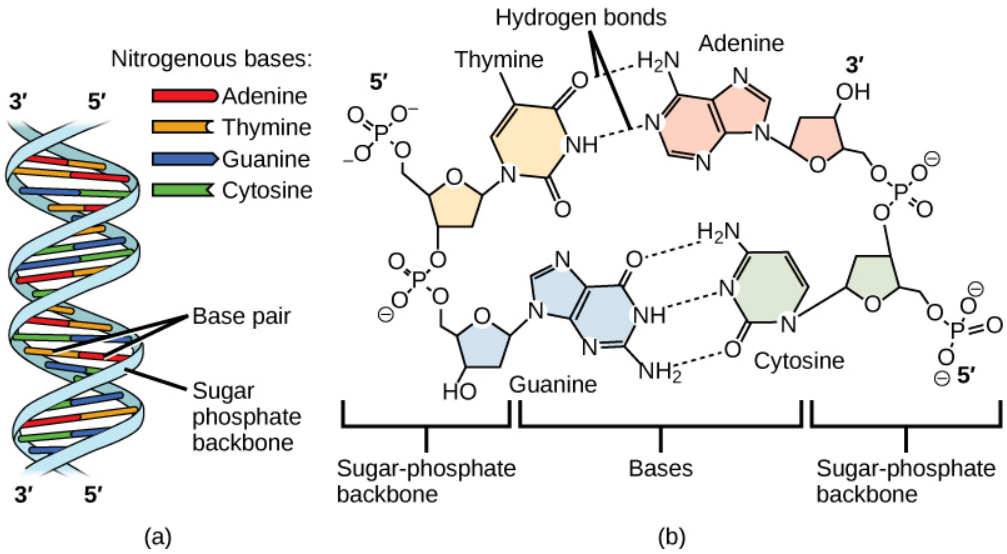


Figure 9.4 DNA (a) forms a double stranded helix, and (b) adenine pairs with thymine and cytosine pairs with guanine. (credit a: modification of work by Jerome Walker, Dennis Myts)

The Structure of RNA

There is a second nucleic acid in all cells called ribonucleic acid, or RNA. Like DNA, RNA is a polymer of nucleotides. Each of the nucleotides in RNA is made up of a nitrogenous base, a five-carbon sugar, and a phosphate group. In the case of RNA, the five-carbon sugar is ribose, not deoxyribose. Ribose has a hydroxyl group at the 2′ carbon, unlike deoxyribose, which has only a hydrogen atom (Figure 9.5).

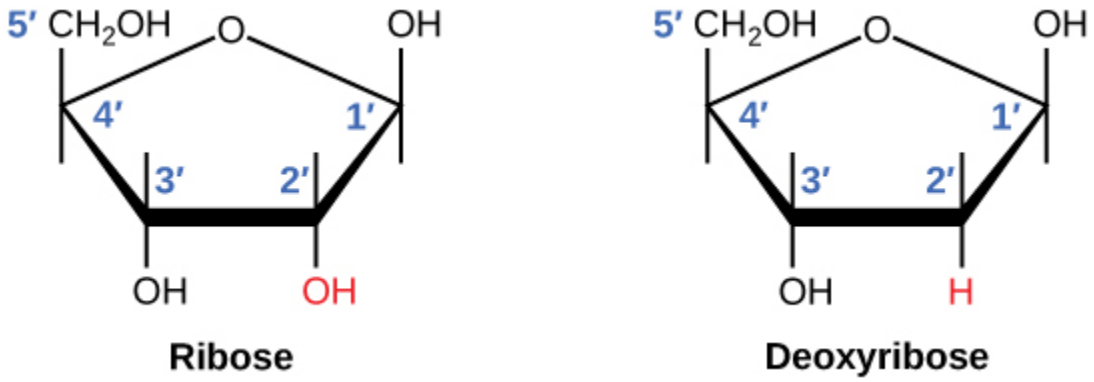


Figure 9.5 The difference between the ribose found in RNA and the deoxyribose found in DNA is that ribose has a hydroxyl group at the 2′ carbon.

RNA nucleotides contain the nitrogenous bases adenine, cytosine, and guanine. However, they do **not contain thymine**, which is instead **replaced by uracil**, symbolized by a “U.” RNA exists as a single-stranded molecule rather than a double-stranded helix. Molecular biologists have named several kinds of RNA on the basis of their function. These include messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA)—molecules that are involved in the production of proteins from the DNA code.

How DNA Is Arranged in the Cell

DNA is a working molecule; it must be replicated when a cell is ready to divide, and it must be "read" to produce the molecules, such as proteins, to carry out the functions of the cell. For this reason, the DNA is protected and packaged in very specific ways. In addition, DNA molecules can be very long. Stretched end-to-end, the DNA molecules in a single human cell would come to a **length of about 2 meters**. Thus, the DNA for a cell must be packaged in a very ordered way to fit and function within a structure (the cell) that is not visible to the naked eye. The chromosomes of prokaryotes are much simpler than those of eukaryotes in many of their features (Figure 9.6). Most prokaryotes contain a single, circular chromosome that is found in an area in the cytoplasm called the nucleoid.

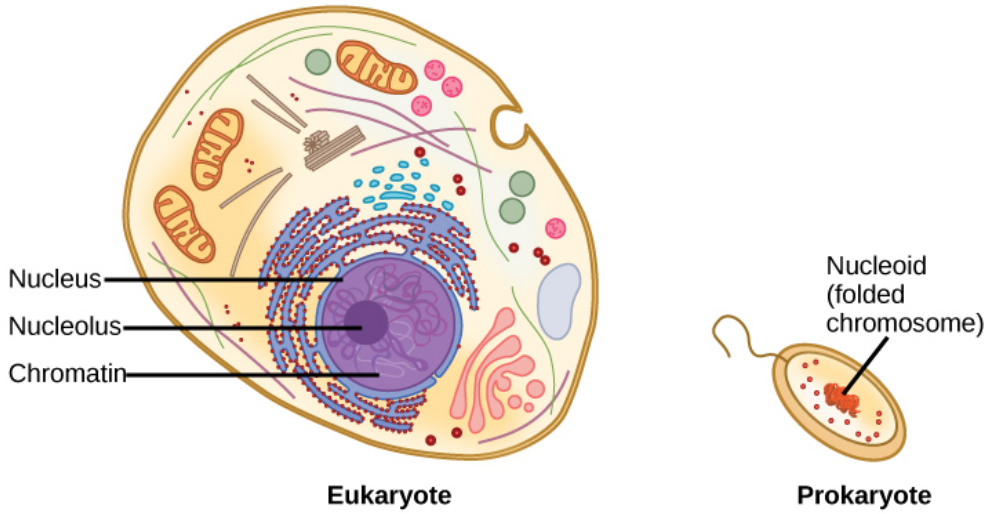


Figure 9.6 A eukaryote contains a well-defined nucleus, whereas in prokaryotes, the chromosome lies in the cytoplasm in an area called the nucleoid.

The size of the genome in one of the most well-studied prokaryotes, *Escherichia coli*, is 4.6 million base pairs, which would extend a distance of about 1.6 mm if stretched out. So how does this fit inside a small bacterial cell? The DNA is twisted beyond the double helix in what is known as supercoiling. Some proteins are known to be involved in the supercoiling; other proteins and enzymes help in maintaining the supercoiled structure.

Eukaryotes, whose chromosomes each consist of a linear DNA molecule, employ a different type of packing strategy to fit their DNA inside the nucleus. At the most basic level, DNA is wrapped around proteins known as histones to form structures called nucleosomes. The DNA is wrapped tightly around the histone core. This nucleosome is linked to the next one by a short strand of DNA that is free of histones. This is also known as the "beads on a string" structure; the nucleosomes are the "beads" and the short lengths of DNA between them are the "string." The nucleosomes, with their DNA coiled around them, stack compactly onto each other to form a 30-nm–wide fiber. This fiber is further coiled into a thicker and more compact structure. At the metaphase stage of mitosis, when the chromosomes are lined up in the center of the cell, the chromosomes are at their most compacted. They are approximately 700 nm in width, and are found in association with scaffold proteins.

In interphase, the phase of the cell cycle between mitoses at which the chromosomes are decondensed,

eukaryotic chromosomes have two distinct regions that can be distinguished by staining. There is a tightly packaged region that stains darkly, and a less dense region. The darkly staining regions usually contain genes that are not active, and are found in the regions of the centromere and telomeres. The lightly staining regions usually contain genes that are active, with DNA packaged around nucleosomes but not further compacted.

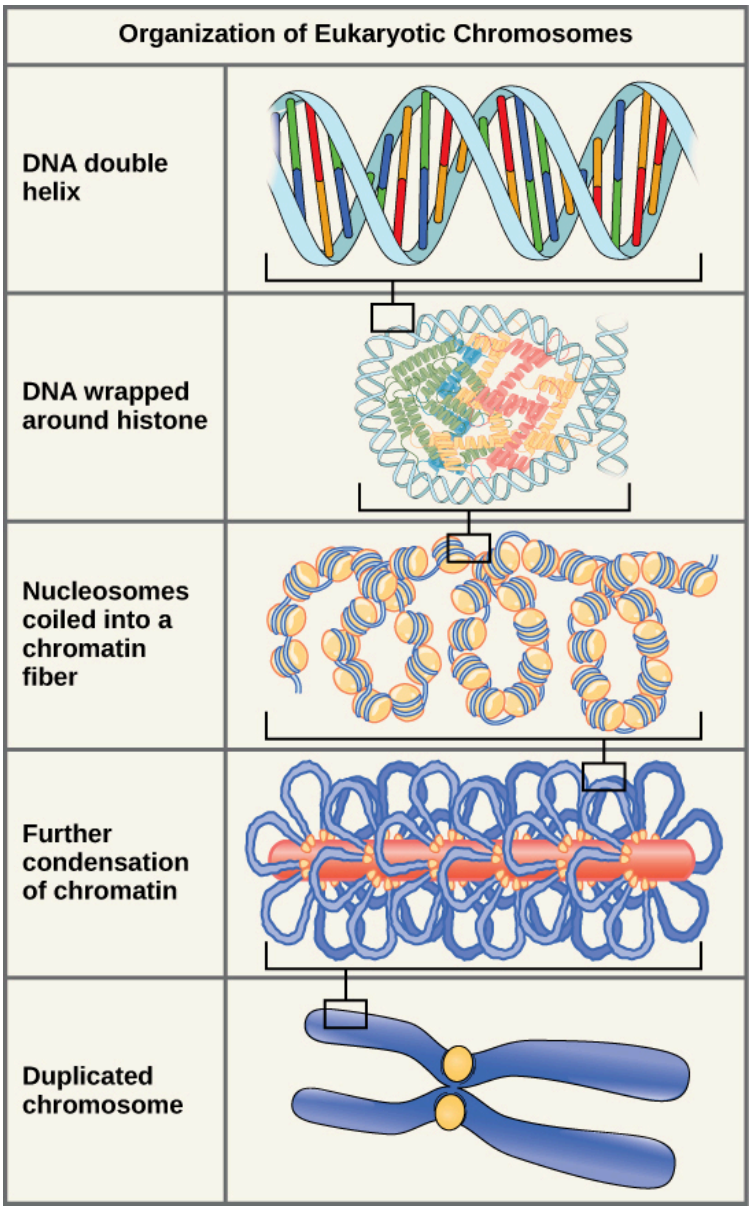


Figure 9.7 These figures illustrate the compaction of the eukaryotic chromosome.

Concept in Action

Watch this animation of DNA packaging.

Section Summary

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4577#h5p-53

The model of the double-helix structure of DNA was proposed by Watson and Crick. The DNA molecule is a polymer of nucleotides. Each nucleotide is composed of a nitrogenous base, a five-carbon sugar (deoxyribose), and a phosphate group. There are four nitrogenous bases in DNA, two purines (adenine and guanine) and two pyrimidines (cytosine and thymine). A DNA molecule is composed of two strands. Each strand is composed of nucleotides bonded together covalently between the phosphate group of one and the deoxyribose sugar of the next. From this backbone extend the bases. The bases of one strand bond to the bases of the second strand with hydrogen bonds. Adenine always bonds with thymine, and cytosine always bonds with guanine. The bonding causes the two strands to spiral around each other in a shape called a double helix. Ribonucleic acid (RNA) is a second nucleic acid found in cells. RNA is a single-stranded polymer of nucleotides. It also differs from DNA in that it contains the sugar ribose, rather than deoxyribose, and the nucleotide uracil rather than thymine. Various RNA molecules function in the process of forming proteins from the genetic code in DNA.

Prokaryotes contain a single, double-stranded circular chromosome. Eukaryotes contain double-stranded linear DNA molecules packaged into chromosomes. The DNA helix is wrapped around proteins to form nucleosomes. The protein coils are further coiled, and during mitosis and meiosis, the chromosomes become even more greatly coiled to facilitate their movement. Chromosomes have two distinct regions which can be distinguished by staining, reflecting different degrees of packaging and determined by whether the DNA in a region is being expressed (euchromatin) or not (heterochromatin).

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4577#h5p-54

Glossary

deoxyribose: a five-carbon sugar molecule with a hydrogen atom rather than a hydroxyl group in the 2′ position; the sugar component of DNA nucleotides

double helix: the molecular shape of DNA in which two strands of nucleotides wind around each other in a spiral shape

nitrogenous base: a nitrogen-containing molecule that acts as a base; often referring to one of the purine or pyrimidine components of nucleic acids

phosphate group: a molecular group consisting of a central phosphorus atom bound to four oxygen atoms

9.2 DNA Replication

Learning Objectives

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 9.8).

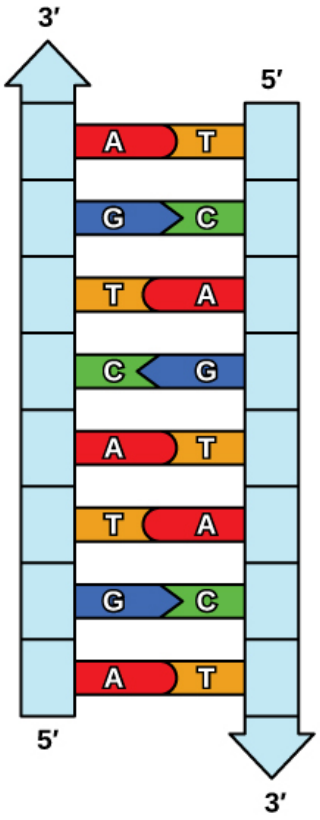


Figure 9.8 The two strands of DNA are complementary, meaning the sequence of bases in one strand can be used to create the correct sequence of bases in the other strand.

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 9.9).

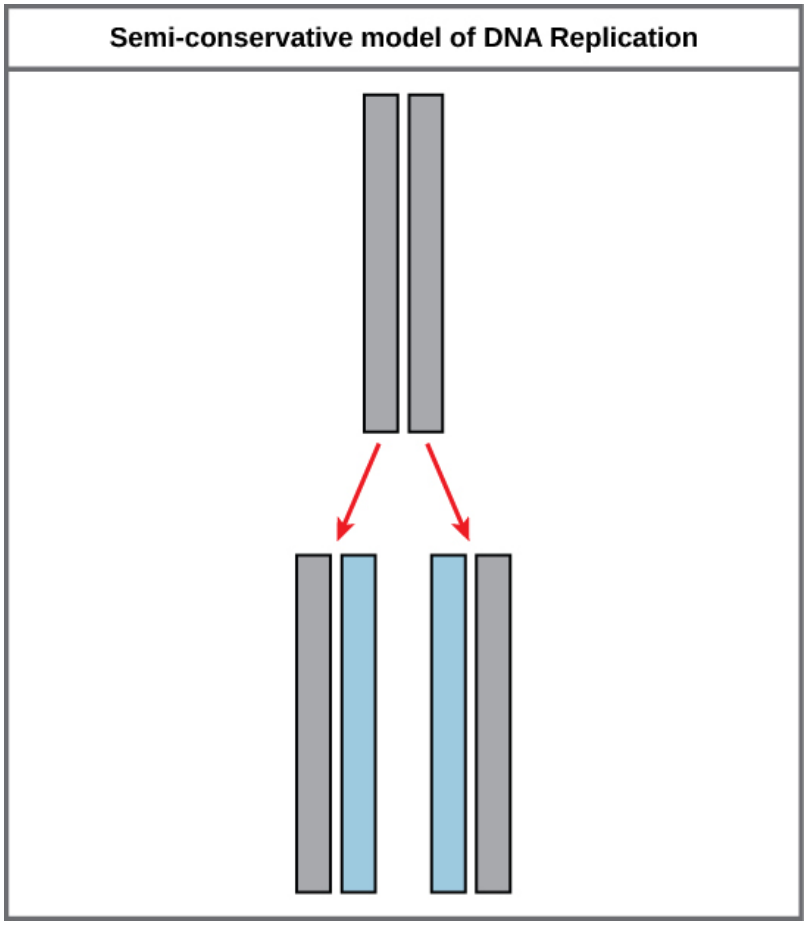


Figure 9.9 The semiconservative model of DNA replication is shown. Gray indicates the original DNA strands, and blue indicates newly synthesized DNA.

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 9.10). 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.

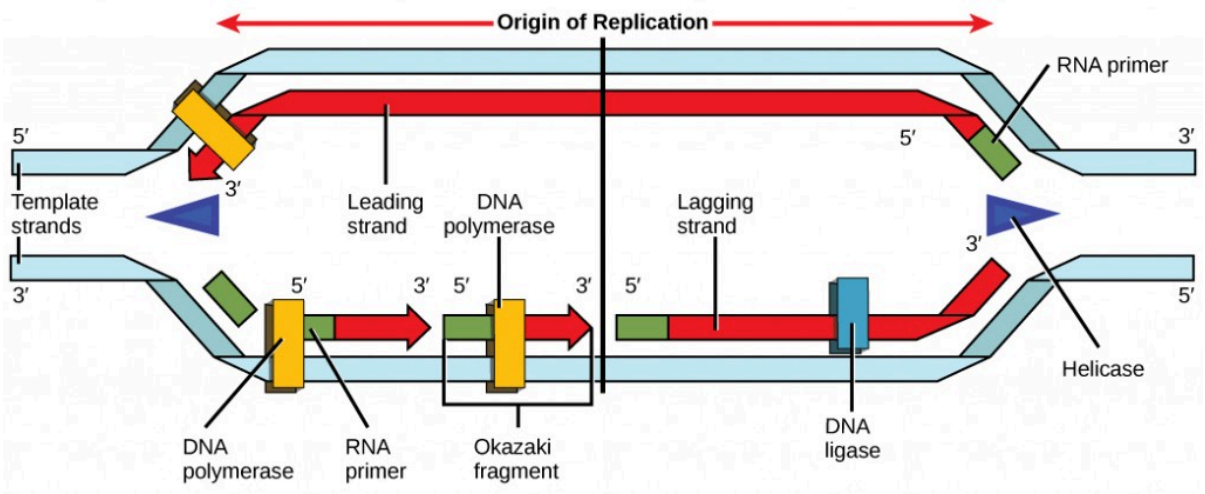


Figure 9.10 A replication fork is formed by the opening of the origin of replication, and helicase separates the DNA strands. An RNA primer is synthesized, and is elongated by the DNA polymerase. On the leading strand, DNA is synthesized continuously, whereas on the lagging strand, DNA is synthesized in short stretches. The DNA fragments are joined by DNA ligase (not shown).

You isolate a cell strain in which the joining together of Okazaki fragments is impaired and suspect that

a mutation has occurred in an enzyme found at the replication fork. Which enzyme is most likely to be mutated?

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 9.11) 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.

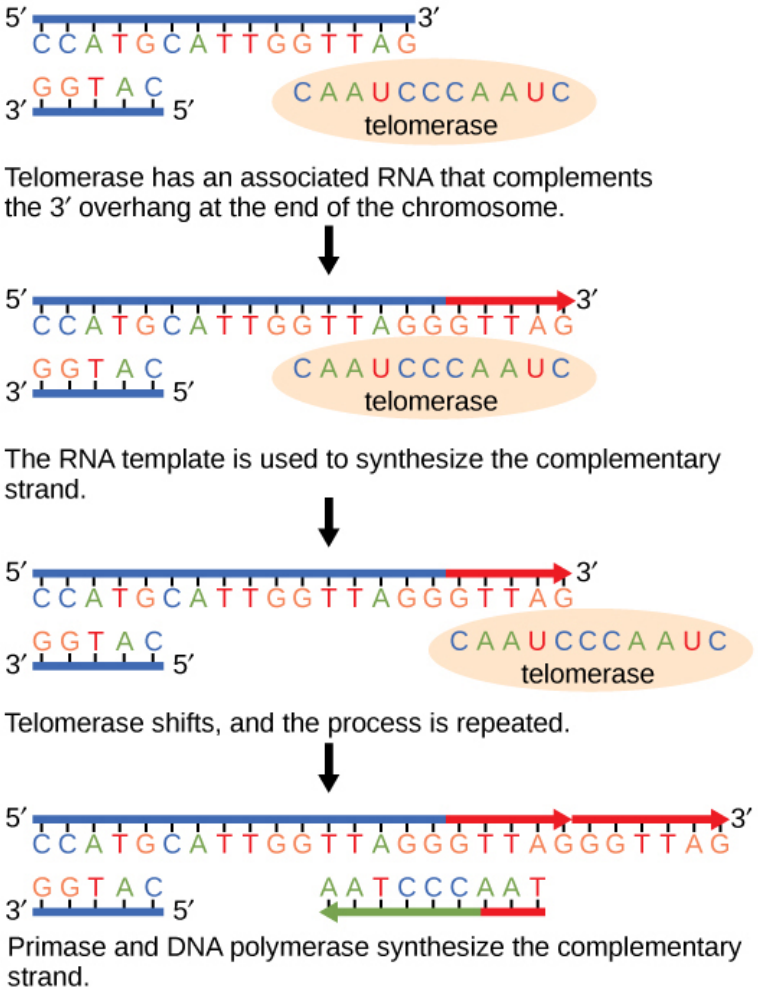


Figure 9.11 The ends of linear chromosomes are maintained by the action of the telomerase enzyme.

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 9.12) received the Nobel Prize for Medicine and Physiology in 2009.

Figure 9.12 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 Replication in Prokaryotes

Recall that the prokaryotic chromosome is a circular molecule with a less extensive coiling structure than eukaryotic chromosomes. The eukaryotic chromosome is linear and highly coiled around proteins. While there are many similarities in the DNA replication process, these structural differences necessitate some differences in the DNA replication process in these two life forms.

DNA replication has been extremely well-studied in prokaryotes, primarily because of the small size of the genome and large number of variants available. *Escherichia coli* has 4.6 million base pairs in a single circular chromosome, and all of it gets replicated in approximately 42 minutes, starting from a single origin of replication and proceeding around the chromosome in both directions. This means that approximately 1000 nucleotides are added per second. The process is much more rapid than in eukaryotes. The table below summarizes the differences between prokaryotic and eukaryotic replications.

Differences between Prokaryotic and Eukaryotic Replications

Property	Prokaryotes	Eukaryotes
Origin of replication	Single	Multiple
Rate of replication	1000 nucleotides/s	50 to 100 nucleotides/s
Chromosome structure	circular	linear
Telomerase	Not present	Present

Concept in Action

Click through a tutorial on DNA replication.

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 9.13 **a**). 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 9.13 **b**). 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 9.13 **c**). 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.

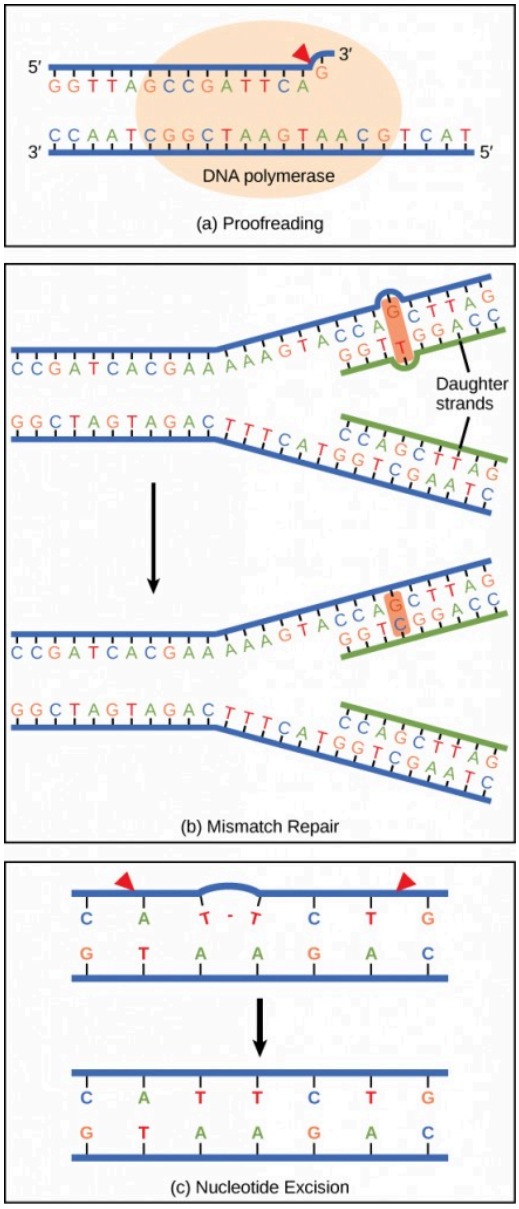


Figure 9.13 Proofreading by DNA polymerase (a) corrects errors during replication. In mismatch repair (b), the incorrectly added base is detected after replication. The mismatch repair proteins detect this base and remove it from the newly synthesized strand by nuclease action. The gap is now filled 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.

Section Summary

DNA replicates by a **semi-conservative** method in which each of the two parental DNA strands act as a template for new DNA to be synthesized. After replication, each DNA has one parental or "old" strand, and one daughter or "new" strand.

Replication in eukaryotes starts at multiple origins of replication, while replication in prokaryotes starts from a single origin of replication. The DNA is opened with enzymes, resulting in the formation of the replication fork. Primase synthesizes an RNA primer to initiate synthesis by DNA polymerase, which can add nucleotides in only one direction. One strand is synthesized continuously in the direction of the replication fork; this is called the leading strand. The other strand is synthesized in a direction away from the replication fork, in short stretches of DNA known as Okazaki fragments. This strand is known as the lagging strand. Once replication is completed, the RNA primers are replaced by DNA nucleotides and the DNA is sealed with DNA ligase.

The ends of eukaryotic chromosomes pose a problem, as polymerase is unable to extend them without a primer. Telomerase, an enzyme with an inbuilt RNA template, extends the ends by copying the RNA template and extending one end of the chromosome. DNA polymerase can then extend the DNA using the primer. In this way, the ends of the chromosomes are protected. Cells have mechanisms for repairing DNA when it becomes damaged or errors are made in replication. These mechanisms include mismatch repair to replace nucleotides that are paired with a non-complementary base and nucleotide excision repair, which removes bases that are damaged such as thymine dimers.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4586#h5p-55

Glossary

DNA ligase: the enzyme that catalyzes the joining of DNA fragments together

DNA polymerase: an enzyme that synthesizes a new strand of DNA complementary to a template strand

helicase: an enzyme that helps to open up the DNA helix during DNA replication by breaking the hydrogen bonds

lagging strand: during replication of the 3′ to 5′ strand, the strand that is replicated in short fragments and away from the replication fork

leading strand: the strand that is synthesized continuously in the 5′ to 3′ direction that is synthesized in the direction of the replication fork

mismatch repair: a form of DNA repair in which non-complementary nucleotides are recognized, excised, and replaced with correct nucleotides

mutation: a permanent variation in the nucleotide sequence of a genome

nucleotide excision repair: a form of DNA repair in which the DNA molecule is unwound and separated in the region of the nucleotide damage, the damaged nucleotides are removed and replaced with new nucleotides using the complementary strand, and the DNA strand is resealed and allowed to rejoin its complement

Okazaki fragments: the DNA fragments that are synthesized in short stretches on the lagging strand
primer: a short stretch of RNA nucleotides that is required to initiate replication and allow DNA polymerase to bind and begin replication

replication fork: the Y-shaped structure formed during the initiation of replication

semiconservative replication: the method used to replicate DNA in which the double-stranded molecule is separated and each strand acts as a template for a new strand to be synthesized, so the resulting DNA molecules are composed of one new strand of nucleotides and one old strand of nucleotides

telomerase: an enzyme that contains a catalytic part and an inbuilt RNA template; it functions to maintain telomeres at chromosome ends

telomere: the DNA at the end of linear chromosomes

Footnotes

1 Mariella Jaskelioff, et al., “Telomerase reactivation reverses tissue degeneration in aged telomerase-deficient mice,” *Nature*, 469 (2011):102–7.

9.3 Transcription

Learning Objectives

By the end of this section, you will be able to:

- Explain the central dogma
- Explain the main steps of transcription
- Describe how eukaryotic mRNA is processed

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 module discusses the details of transcription.

The Central Dogma: 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** (Figure 9.14), which states that genes specify the sequences of mRNAs, which in turn specify the sequences of proteins.

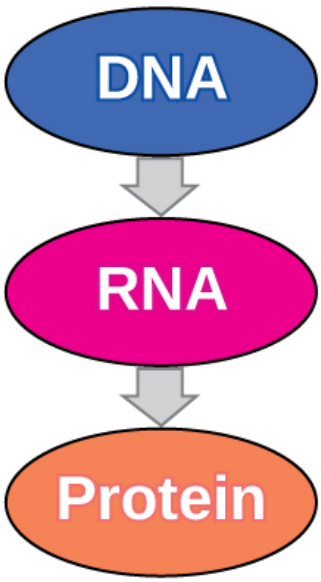


Figure 9.14 The central dogma states that DNA encodes RNA, which in turn encodes protein.

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 module, 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 9.15).

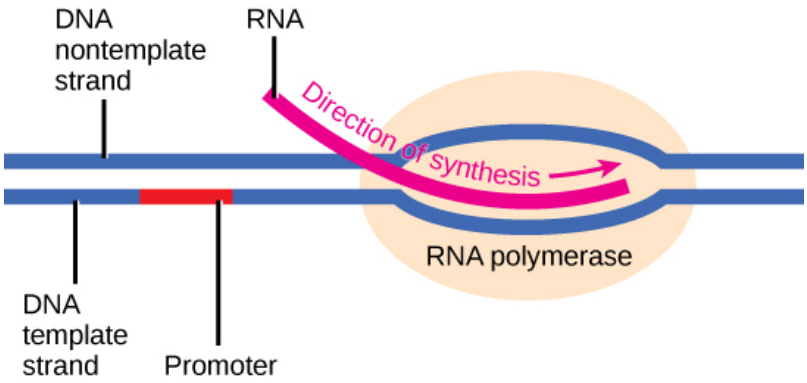


Figure 9.15 The initiation of transcription begins when DNA is unwound, forming a transcription bubble. Enzymes and other proteins involved in transcription bind at the promoter.

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 9.16).

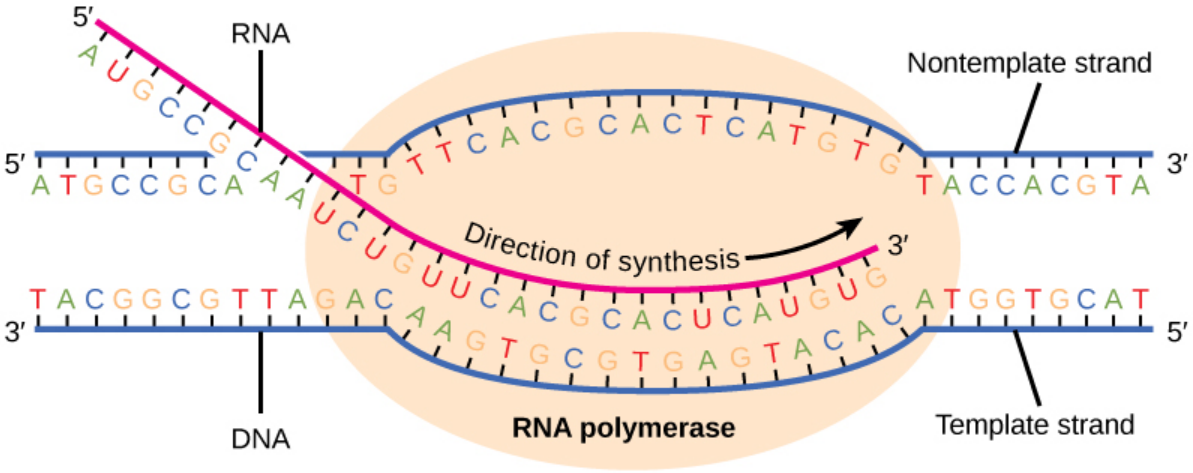


Figure 9.16 During elongation, RNA polymerase tracks along the DNA template, synthesizes mRNA in the 5′ to 3′ direction, and unwinds then rewinds the DNA as it is read.

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 9.17). In contrast, the presence of a nucleus in eukaryotic cells precludes simultaneous transcription and translation.

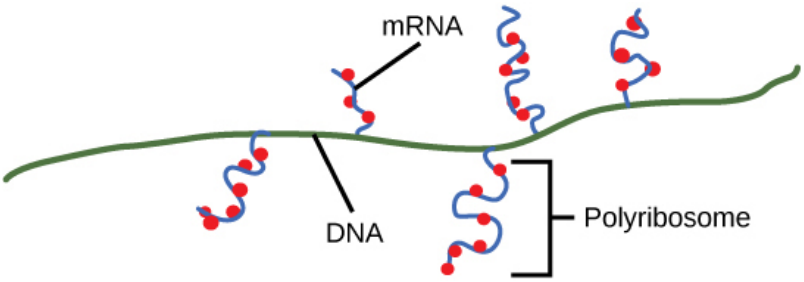


Figure 9.17 Multiple polymerases can transcribe a single bacterial gene while numerous ribosomes concurrently translate the mRNA transcripts into polypeptides. In this way, a specific protein can rapidly reach a high concentration in the bacterial cell.

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 9.18). Introns are removed and degraded while the pre-mRNA is still in the nucleus.

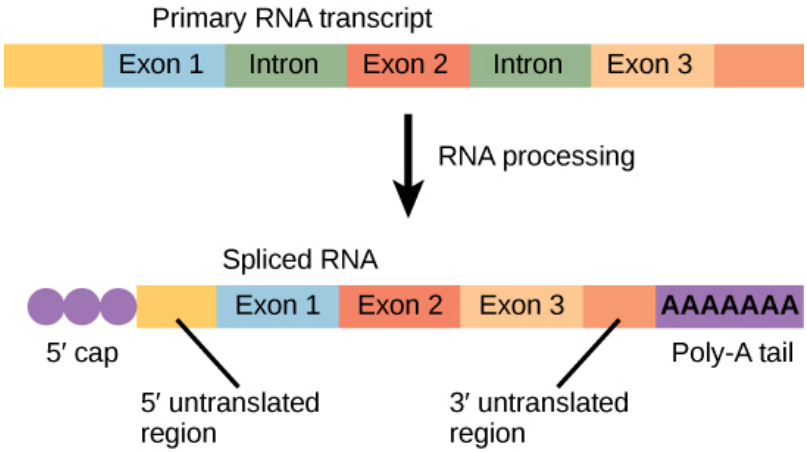


Figure 9.18 Eukaryotic mRNA contains introns that must be spliced out. A 5′ cap and 3′ tail are also added.

Section Summary

In prokaryotes, mRNA synthesis is initiated at a promoter sequence on the DNA template. Elongation synthesizes new mRNA. Termination liberates the mRNA and occurs by mechanisms that stall the RNA polymerase and cause it to fall off the DNA template. Newly transcribed eukaryotic mRNAs are modified with a cap and a poly-A tail. These structures protect the mature mRNA from degradation and help export it from the nucleus. Eukaryotic mRNAs also undergo splicing, in which introns are removed and exons are reconnected with single-nucleotide accuracy. Only finished mRNAs are exported from the nucleus to the cytoplasm.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4593#h5p-56

Glossary

exon: a sequence present in protein-coding mRNA after completion of pre-mRNA splicing

intron: non–protein-coding intervening sequences that are spliced from mRNA during processing

mRNA: messenger RNA; a form of RNA that carries the nucleotide sequence code for a protein sequence that is translated into a polypeptide sequence

nontemplate strand: the strand of DNA that is not used to transcribe mRNA; this strand is identical to the mRNA except that T nucleotides in the DNA are replaced by U nucleotides in the mRNA

promoter: a sequence on DNA to which RNA polymerase and associated factors bind and initiate transcription

RNA polymerase: an enzyme that synthesizes an RNA strand from a DNA template strand

splicing: the process of removing introns and reconnecting exons in a pre-mRNA

template strand: the strand of DNA that specifies the complementary mRNA molecule

transcription bubble: the region of locally unwound DNA that allows for transcription of mRNA

9.4 Translation

Learning Objectives

By the end of this section, you will be able to:

- Describe the different steps in protein synthesis
- Discuss the role of ribosomes in protein synthesis
- Describe the genetic code and how the nucleotide sequence determines the amino acid and the protein sequence

The synthesis of proteins is one of a cell's most energy-consuming metabolic processes. In turn, 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**. 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 9.19).

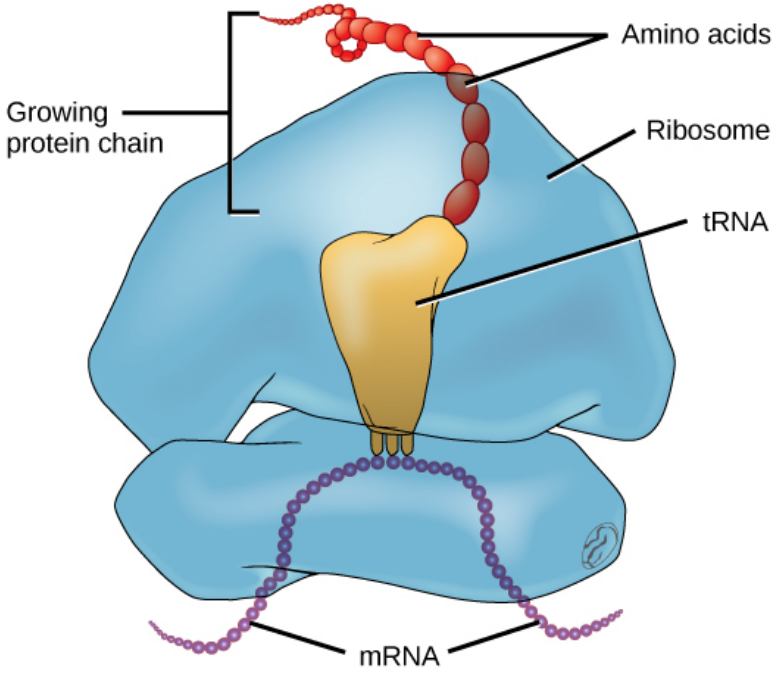


Figure 9.19 The protein synthesis machinery includes the large and small subunits of the ribosome, mRNA, and tRNA. (credit: modification of work by NIGMS, NIH)

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 in 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

Universality of the Genetic Code

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4599#h5p-57

Mutations

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4599#h5p-58

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** ($4 \times 4 \times 4$) possible combinations; therefore, a given amino acid is encoded by more than one nucleotide triplet (Figure 9.20).

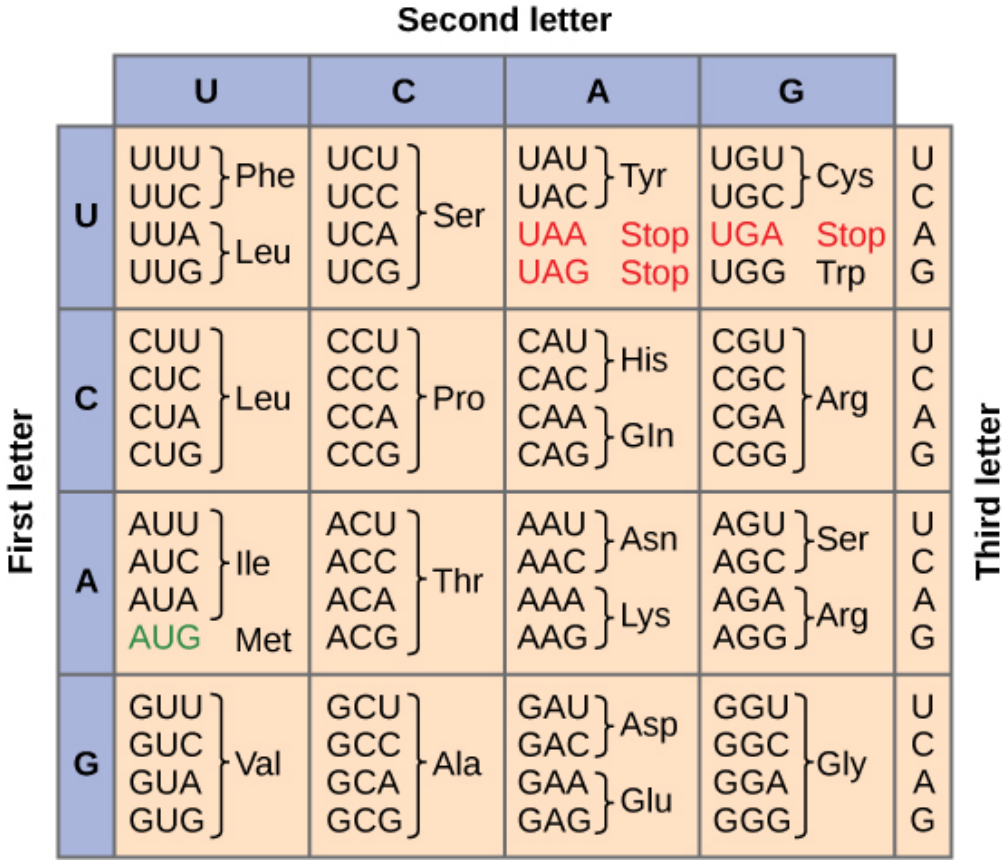


Figure 9.20 This figure shows the genetic code for translating each nucleotide triplet, or codon, in mRNA into an amino acid or a termination signal in a nascent protein. (credit: modification of work by NIH)

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 9.21). 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.

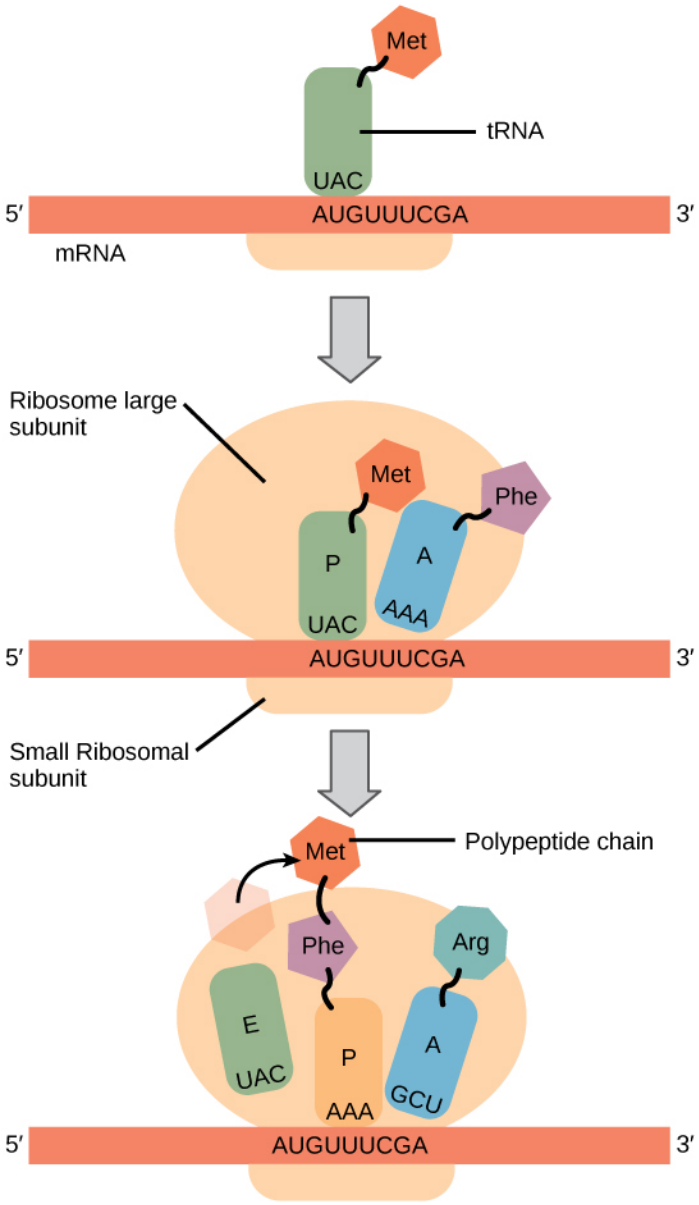


Figure 9.21 Translation begins when a tRNA anticodon recognizes a codon on the mRNA. The large ribosomal subunit joins the small subunit, and a second tRNA is recruited. As the mRNA moves relative to the ribosome, the polypeptide chain is formed. Entry of a release factor into the A site terminates translation 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.

Concept in Action

Transcribe a gene and translate it to protein using complementary pairing and the genetic code at this site.

Section Summary

The central dogma describes the flow of genetic information in the cell from genes to mRNA to proteins. Genes are used to make mRNA by the process of transcription; mRNA is used to synthesize proteins by the process of translation. The genetic code is the correspondence between the three-nucleotide mRNA codon and an amino acid. The genetic code is "translated" by the tRNA molecules, which associate a specific codon with a specific amino acid. The genetic code is degenerate because 64 triplet codons in mRNA specify only 20 amino acids and three stop codons. This means that more than one codon corresponds to an amino acid. Almost every species on the planet uses the same genetic code.

The players in translation include the mRNA template, ribosomes, tRNAs, and various enzymatic factors. The small ribosomal subunit binds to the mRNA template. Translation begins at the initiating AUG on the mRNA. The formation of bonds occurs between sequential amino acids specified by the mRNA template according to the genetic code. The ribosome accepts charged tRNAs, and as it steps along the mRNA, it catalyzes bonding between the new amino acid and the end of the growing polypeptide. The entire mRNA is translated in three-nucleotide "steps" of the ribosome. When a stop codon is encountered, a release factor binds and dissociates the components and frees the new protein.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4599#h5p-59

Glossary

codon: three consecutive nucleotides in mRNA that specify the addition of a specific amino acid or the release of a polypeptide chain during translation

genetic code: the amino acids that correspond to three-nucleotide codons of mRNA

rRNA: ribosomal RNA; molecules of RNA that combine to form part of the ribosome

stop codon: one of the three mRNA codons that specifies termination of translation

start codon: the AUG (or, rarely GUG) on an mRNA from which translation begins; always specifies methionine

tRNA: transfer RNA; an RNA molecule that contains a specific three-nucleotide anticodon sequence to pair with the mRNA codon and also binds to a specific amino acid

9.5 How Genes Are Regulated

Learning Objectives

By the end of this section, you will be able to:

- Discuss why every cell does not express all of its genes
- Describe how prokaryotic gene expression occurs at the transcriptional level
- Understand that eukaryotic gene expression occurs at the epigenetic, transcriptional, post-transcriptional, translational, and post-translational levels

For a cell to function properly, necessary proteins must be synthesized at the proper time. All organisms and cells control or regulate the transcription and translation of their DNA into protein. The process of turning on a gene to produce RNA and protein is called gene expression. Whether in a simple unicellular organism or in a complex multicellular organism, each cell controls when and how its genes are expressed. For this to occur, there must be a mechanism to control when a gene is expressed to make RNA and protein, how much of the protein is made, and when it is time to stop making that protein because it is no longer needed.

Cells in multicellular organisms are specialized; cells in different tissues look very different and perform different functions. For example, a muscle cell is very different from a liver cell, which is very different from a skin cell. These differences are a consequence of the expression of different sets of genes in each of these cells. All cells have certain basic functions they must perform for themselves, such as converting the energy in sugar molecules into energy in ATP. Each cell also has many genes that are not expressed, and expresses many that are not expressed by other cells, such that it can carry out its specialized functions. In addition, cells will turn on or off certain genes at different times in response to changes in the environment or at different times during the development of the organism. Unicellular organisms, both eukaryotic and prokaryotic, also turn on and off genes in response to the demands of their environment so that they can respond to special conditions.

The control of gene expression is extremely complex. Malfunctions in this process are detrimental to the cell and can lead to the development of many diseases, including cancer.

Prokaryotic versus Eukaryotic Gene Expression

To understand how gene expression is regulated, we must first understand how a gene becomes a functional protein in a cell. The process occurs in both prokaryotic and eukaryotic cells, just in slightly different fashions.

Because prokaryotic organisms lack a cell nucleus, the processes of transcription and translation occur

almost simultaneously. When the protein is no longer needed, transcription stops. As a result, the primary method to control what type and how much protein is expressed in a prokaryotic cell is through the regulation of DNA transcription into RNA. All the subsequent steps happen automatically. When more protein is required, more transcription occurs. Therefore, in prokaryotic cells, the control of gene expression is almost entirely at the transcriptional level.

The first example of such control was discovered using *E. coli* in the 1950s and 1960s by French researchers and is called the *lac* operon. The *lac* operon is a stretch of DNA with three adjacent genes that code for proteins that participate in the absorption and metabolism of lactose, a food source for *E. coli*. When lactose is not present in the bacterium's environment, the *lac* genes are transcribed in small amounts. When lactose is present, the genes are transcribed and the bacterium is able to use the lactose as a food source. The operon also contains a promoter sequence to which the RNA polymerase binds to begin transcription; between the promoter and the three genes is a region called the operator. When there is no lactose present, a protein known as a repressor binds to the operator and prevents RNA polymerase from binding to the promoter, except in rare cases. Thus very little of the protein products of the three genes is made. When lactose is present, an end product of lactose metabolism binds to the repressor protein and prevents it from binding to the operator. This allows RNA polymerase to bind to the promoter and freely transcribe the three genes, allowing the organism to metabolize the lactose.

Eukaryotic cells, in contrast, have intracellular organelles and are much more complex. Recall that in eukaryotic cells, the DNA is contained inside the cell's nucleus and it is transcribed into mRNA there. The newly synthesized mRNA is then transported out of the nucleus into the cytoplasm, where ribosomes translate the mRNA into protein. The processes of transcription and translation are physically separated by the nuclear membrane; transcription occurs only within the nucleus, and translation only occurs outside the nucleus in the cytoplasm. The regulation of gene expression can occur at all stages of the process (Figure 9.22). Regulation may occur when the DNA is uncoiled and loosened from nucleosomes to bind transcription factors (epigenetic level), when the RNA is transcribed (transcriptional level), when RNA is processed and exported to the cytoplasm after it is transcribed (post-transcriptional level), when the RNA is translated into protein (translational level), or after the protein has been made (post-translational level).

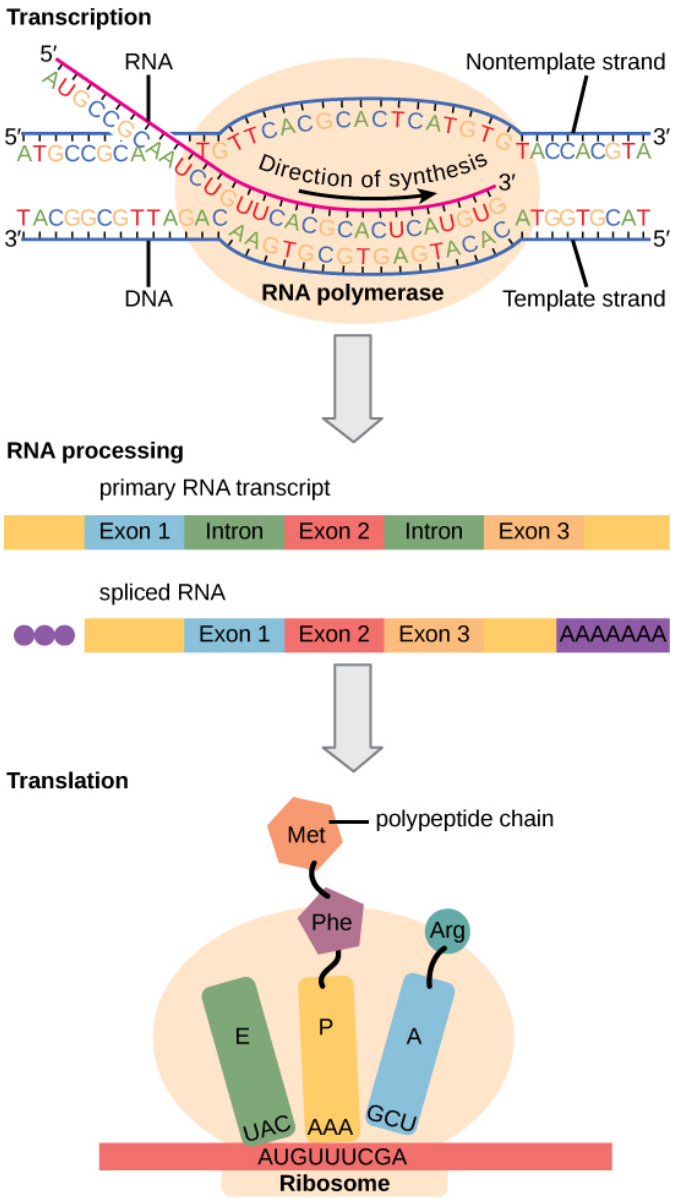


Figure 9.22 Eukaryotic gene expression is regulated during transcription and RNA processing, which take place in the nucleus, as well as during protein translation, which takes place in the cytoplasm. Further regulation may occur through post-translational modifications of proteins.

The differences in the regulation of gene expression between prokaryotes and eukaryotes are summarized in the table below.

Differences in the Regulation of Gene Expression of Prokaryotic and Eukaryotic Organisms

Prokaryotic organisms	Eukaryotic organisms
Lack nucleus	Contain nucleus
RNA transcription and protein translation occur almost simultaneously	• RNA transcription occurs prior to protein translation, and it takes place in the nucleus. RNA translation to protein occurs in the cytoplasm. • RNA post-processing includes addition of a 5′ cap, poly-A tail, and excision of introns and splicing of exons.
Gene expression is regulated primarily at the transcriptional level	Gene expression is regulated at many levels (epigenetic, transcriptional, post-transcriptional, translational, and post-translational)

Alternative RNA Splicing

In the 1970s, genes were first observed that exhibited alternative RNA splicing. Alternative RNA splicing is a mechanism that allows different protein products to be produced from one gene when different combinations of introns (and sometimes exons) are removed from the transcript (Figure 9.23). This alternative splicing can be haphazard, but more often it is controlled and acts as a mechanism of gene regulation, with the frequency of different splicing alternatives controlled by the cell as a way to control the production of different protein products in different cells, or at different stages of development. Alternative splicing is now understood to be a common mechanism of gene regulation in eukaryotes; according to one estimate, 70% of genes in humans are expressed as multiple proteins through alternative splicing.

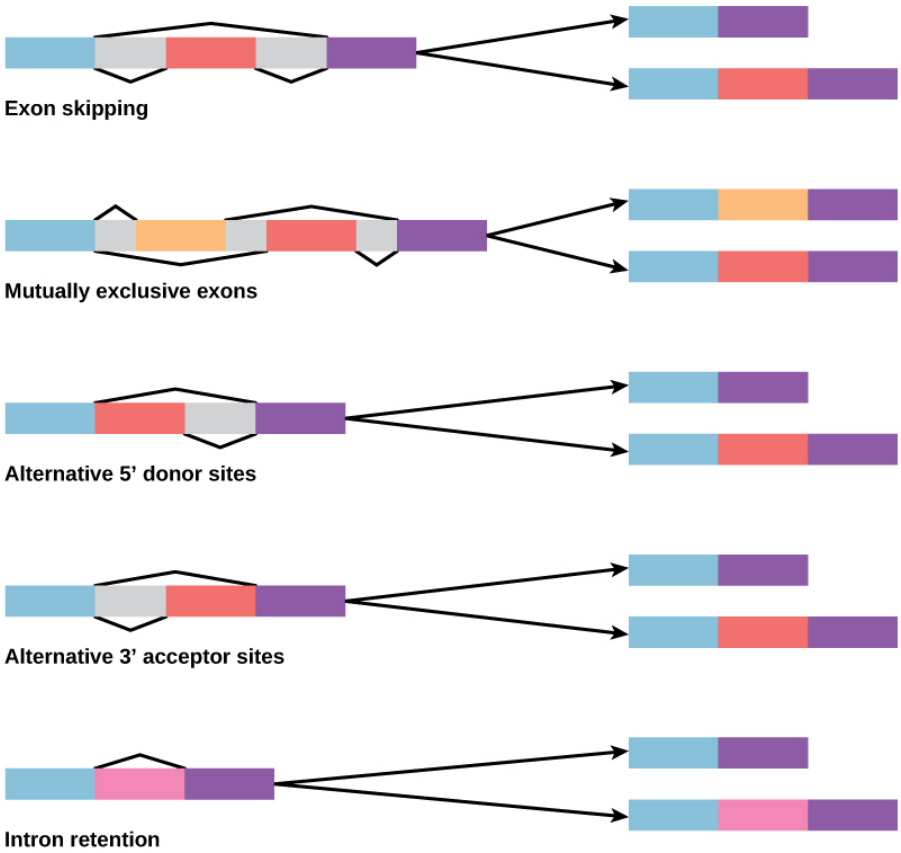


Figure 9.23 There are five basic modes of alternative splicing. Segments of pre-mRNA with exons shown in blue, red, orange, and pink can be spliced to produce a variety of new mature mRNA segments.

How could alternative splicing evolve? Introns have a beginning and ending recognition sequence, and it is easy to imagine the failure of the splicing mechanism to identify the end of an intron and find the end of the next intron, thus removing two introns and the intervening exon. In fact, there are mechanisms in place to prevent such exon skipping, but mutations are likely to lead to their failure. Such "mistakes" would more than likely produce a nonfunctional protein. Indeed, the cause of many genetic diseases is alternative splicing rather than mutations in a sequence. However, alternative splicing would create a protein variant without the loss of the original protein, opening up possibilities for adaptation of the new variant to new functions. Gene duplication has played an important role in the evolution of new functions in a similar way—by providing genes that may evolve without eliminating the original functional protein.

Section Summary

While all somatic cells within an organism contain the same DNA, not all cells within that organism express the same proteins. Prokaryotic organisms express the entire DNA they encode in every cell, but not necessarily all at the same time. Proteins are expressed only when they are needed. Eukaryotic organisms express a subset of the DNA that is encoded in any given cell. In each cell type, the type and amount of protein is regulated by controlling gene expression. To express a protein, the DNA is first transcribed into RNA, which is then translated into proteins. In prokaryotic cells, these processes occur

almost simultaneously. In eukaryotic cells, transcription occurs in the nucleus and is separate from the translation that occurs in the cytoplasm. Gene expression in prokaryotes is regulated only at the transcriptional level, whereas in eukaryotic cells, gene expression is regulated at the epigenetic, transcriptional, post-transcriptional, translational, and post-translational levels.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4603#h5p-60

Glossary

alternative RNA splicing: a post-transcriptional gene regulation mechanism in eukaryotes in which multiple protein products are produced by a single gene through alternative splicing combinations of the RNA transcript

epigenetic: describing non-genetic regulatory factors, such as changes in modifications to histone proteins and DNA that control accessibility to genes in chromosomes

gene expression: processes that control whether a gene is expressed

post-transcriptional: control of gene expression after the RNA molecule has been created but before it is translated into protein

post-translational: control of gene expression after a protein has been created

Chapter 10: Introduction to Biotechnology

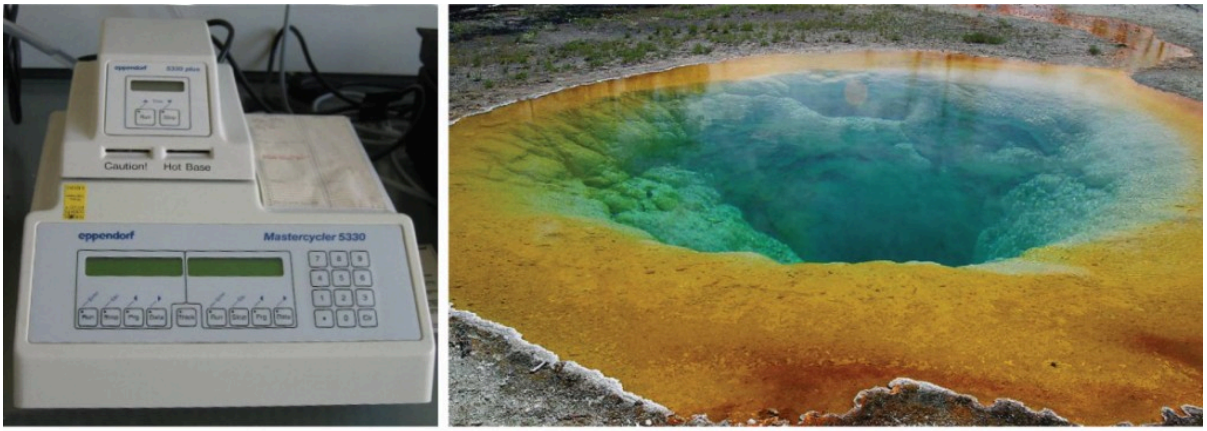


Figure 10.1 (a) A thermal cycler, such as the one shown here, is a basic tool used to study DNA in a process called the polymerase chain reaction (PCR). The polymerase enzyme most often used with PCR comes from a strain of bacteria that lives in (b) the hot springs of Yellowstone National Park. (credit a: modification of work by Magnus Manske; credit b: modification of work by Jon Sullivan)

The latter half of the twentieth century began with the discovery of the structure of DNA, then progressed to the development of the basic tools used to study and **manipulate DNA**. These advances, as well as advances in our understanding of and ability to manipulate cells, have led some to refer to the twenty-first century as the biotechnology century. The rate of discovery and of the development of new applications in medicine, agriculture, and energy is expected to accelerate, bringing huge benefits to humankind and perhaps also significant risks. Many of these developments are expected to raise significant ethical and social questions that human societies have not yet had to consider.

Search for Key Points in Chapter 10

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4606#h5p-61

10.1 Cloning and Genetic Engineering

Learning Objectives

By the end of this section, you will be able to:

- Explain the basic techniques used to manipulate genetic material
- Explain molecular and reproductive cloning

Biotechnology is the use of artificial methods to **modify the genetic material** of living organisms or cells to produce novel compounds or to perform new functions. Biotechnology has been used for improving livestock and crops since the beginning of agriculture through selective breeding. Since the discovery of the structure of DNA in 1953, and particularly since the development of tools and methods to manipulate DNA in the 1970s, biotechnology has become synonymous with the manipulation of organisms' DNA at the molecular level. The primary applications of this technology are in medicine (for the production of vaccines and antibiotics) and in agriculture (for the genetic modification of crops). Biotechnology also has many industrial applications, such as fermentation, the treatment of oil spills, and the production of biofuels, as well as many household applications such as the use of enzymes in laundry detergent.

Manipulating Genetic Material

To accomplish the applications described above, biotechnologists must be able to extract, manipulate, and analyze nucleic acids.

Review of Nucleic Acid Structure

To understand the basic techniques used to work with nucleic acids, remember that nucleic acids are macromolecules made of nucleotides (a sugar, a phosphate, and a nitrogenous base). The phosphate groups on these molecules each have a net negative charge. An entire set of DNA molecules in the nucleus of eukaryotic organisms is called the genome. DNA has two complementary strands linked by hydrogen bonds between the paired bases.

Unlike DNA in eukaryotic cells, RNA molecules leave the nucleus. Messenger RNA (mRNA) is analyzed most frequently because it represents the protein-coding genes that are being expressed in the cell.

Isolation of Nucleic Acids

To study or manipulate nucleic acids, the DNA must first be extracted from cells. Various techniques are used to extract different types of DNA (Figure 10.2). Most nucleic acid extraction techniques involve steps to break open the cell, and then the use of enzymatic reactions to destroy all undesired macromolecules. Cells are broken open using a detergent solution containing buffering compounds. To prevent degradation and contamination, macromolecules such as proteins and RNA are inactivated using enzymes. The DNA is then brought out of solution using alcohol. The resulting DNA, because it is made up of long polymers, forms a gelatinous mass.

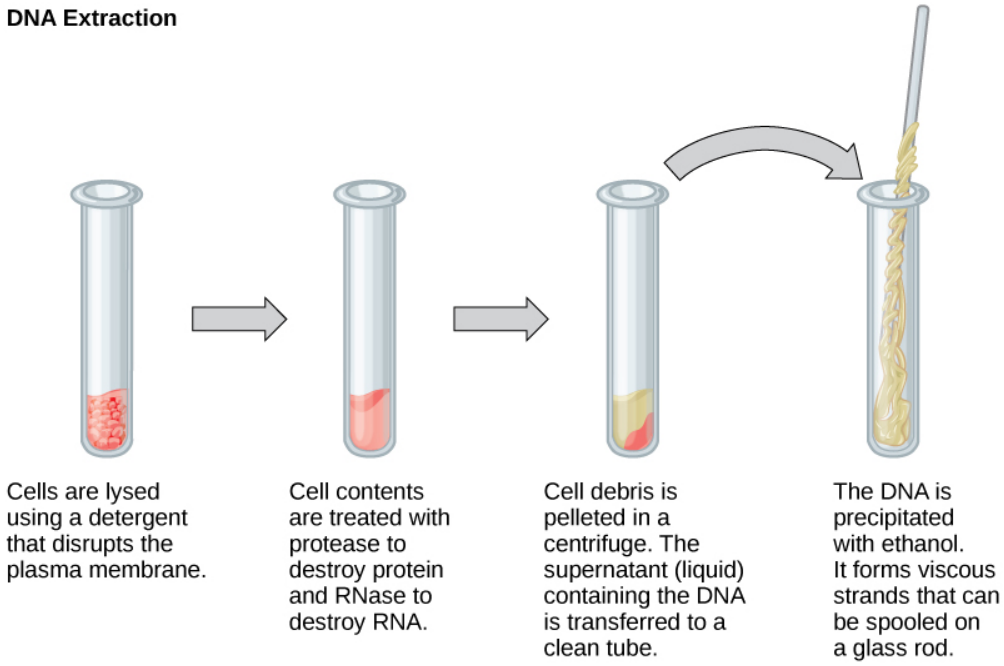


Figure 10.2 This diagram shows the basic method used for the extraction of DNA.

RNA is studied to understand gene expression patterns in cells. RNA is naturally very unstable because enzymes that break down RNA are commonly present in nature. Some are even secreted by our own skin and are very difficult to inactivate. Similar to DNA extraction, RNA extraction involves the use of various buffers and enzymes to inactivate other macromolecules and preserve only the RNA.

Gel Electrophoresis

Because nucleic acids are negatively charged ions at neutral or alkaline pH in an aqueous environment, they can be moved by an electric field. Gel electrophoresis is a technique used to separate charged molecules on the basis of size and charge. The nucleic acids can be separated as whole chromosomes or as fragments. The nucleic acids are loaded into a slot at one end of a gel matrix, an electric current is applied, and negatively charged molecules are pulled toward the opposite end of the gel (the end with the positive electrode). Smaller molecules move through the pores in the gel faster than larger molecules; this difference in the rate of migration separates the fragments on the basis of size. The nucleic acids in a gel matrix are invisible until they are stained with a compound that allows them to be seen, such as a dye. Distinct fragments of nucleic acids appear as bands at specific distances from the

top of the gel (the negative electrode end) that are based on their size (Figure 10.3). A mixture of many fragments of varying sizes appear as a long smear, whereas uncut genomic DNA is usually too large to run through the gel and forms a single large band at the top of the gel.

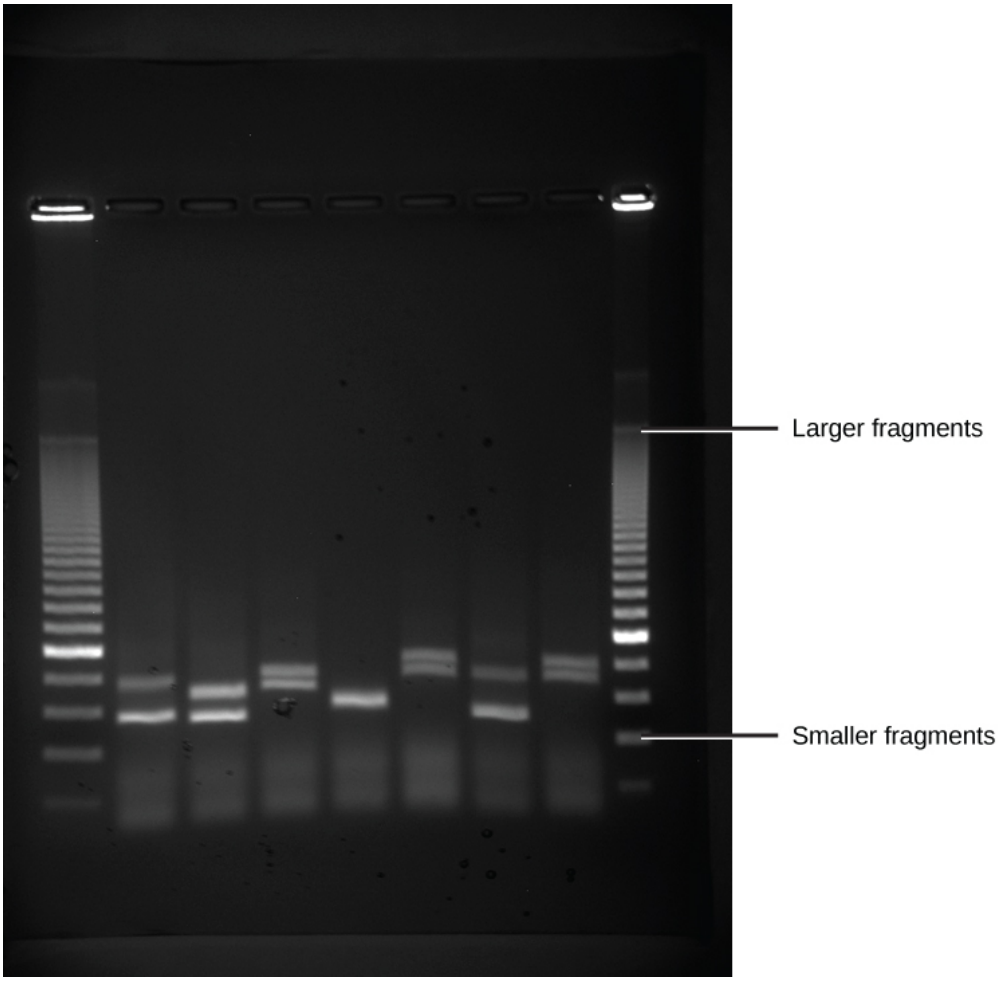


Figure 10.3 Shown are DNA fragments from six samples run on a gel, stained with a fluorescent dye and viewed under UV light. (credit: modification of work by James Jacob, Tompkins Cortland Community College)

Polymerase Chain Reaction

DNA analysis often requires focusing on one or more specific regions of the genome. It also frequently involves situations in which only one or a few copies of a DNA molecule are available for further analysis. These amounts are insufficient for most procedures, such as gel electrophoresis. **Polymerase chain reaction (PCR)** is a technique used to **rapidly increase the number of copies** of specific regions of DNA for further analyses (Figure 10.4). PCR uses a special form of DNA polymerase, the enzyme that replicates DNA, and other short nucleotide sequences called primers that base pair to a specific portion of the DNA being replicated. PCR is used for many purposes in laboratories. These include: 1) the identification of the owner of a DNA sample left at a crime scene; 2) paternity analysis; 3) the comparison of small amounts of ancient DNA with modern organisms; and 4) determining the sequence of nucleotides in a specific region.

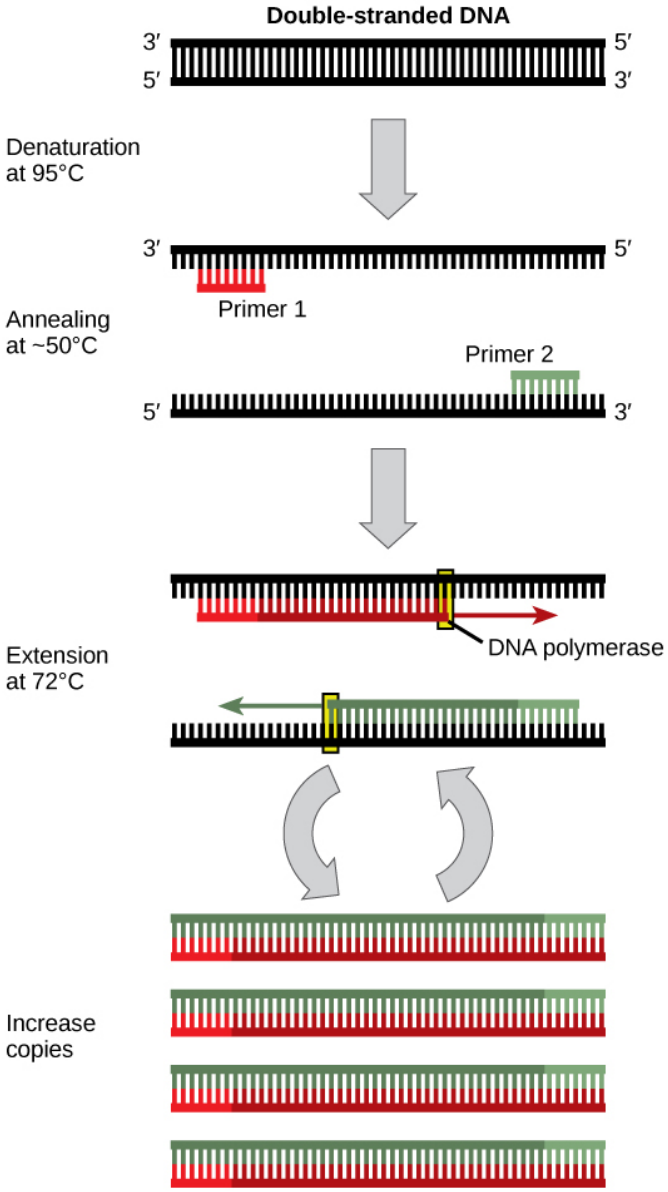


Figure 10.4 Polymerase chain reaction, or PCR, is used to produce many copies of a specific sequence of DNA using a special form of DNA polymerase. Figure showing PCR in 4 steps. First, the double strand of DNA is denatured at 95 degrees Celsius to separate the strands. The 2 strands are then annealed at approximately 50 degrees Celsius using primers. DNA polymerase then extends the new strands at 72 degrees Celsius. The fourth step shows that this procedure takes place many times, resulting in an increase in copies of the original DNA.

Cloning

In general, cloning means the creation of a perfect replica. Typically, the word is used to describe the creation of a genetically identical copy. In biology, the re-creation of a whole organism is referred to as

"reproductive cloning." Long before attempts were made to clone an entire organism, researchers learned how to copy short stretches of DNA—a process that is referred to as molecular cloning.

Molecular Cloning

Cloning allows for the creation of **multiple copies of genes, expression of genes, and study of specific genes**. To get the DNA fragment into a bacterial cell in a form that will be copied or expressed, the fragment is first inserted into a plasmid. A **plasmid** (also called a vector in this context) is a small circular DNA molecule that replicates independently of the chromosomal DNA in bacteria. In cloning, the plasmid molecules can be used to provide a "vehicle" in which to insert a desired DNA fragment. Modified plasmids are usually reintroduced into a bacterial host for replication. As the bacteria divide, they copy their own DNA (including the plasmids). The inserted DNA fragment is copied along with the rest of the bacterial DNA. In a bacterial cell, the fragment of DNA from the human genome (or another organism that is being studied) is referred to as foreign DNA to differentiate it from the DNA of the bacterium (the host DNA).

Plasmids occur naturally in bacterial populations (such as *Escherichia coli*) and have genes that can contribute favorable traits to the organism, such as antibiotic resistance (the ability to be unaffected by antibiotics). Plasmids have been highly engineered as vectors for molecular cloning and for the subsequent large-scale production of important molecules, such as insulin. A valuable characteristic of plasmid vectors is the ease with which a foreign DNA fragment can be introduced. These plasmid vectors contain many short DNA sequences that can be cut with different commonly available restriction enzymes. **Restriction enzymes** (also called restriction endonucleases) recognize specific DNA sequences and cut them in a predictable manner; they are naturally produced by bacteria as a defense mechanism against foreign DNA. Many restriction enzymes make **staggered cuts** in the two strands of DNA, such that the cut ends have a 2- to 4-nucleotide single-stranded overhang. The sequence that is recognized by the restriction enzyme is a four- to eight-nucleotide sequence that is a **palindrome**. Like with a word palindrome, this means the sequence reads the same forward and backward. In most cases, the sequence reads the same forward on one strand and backward on the complementary strand. When a staggered cut is made in a sequence like this, the overhangs are complementary (Figure 10.5).

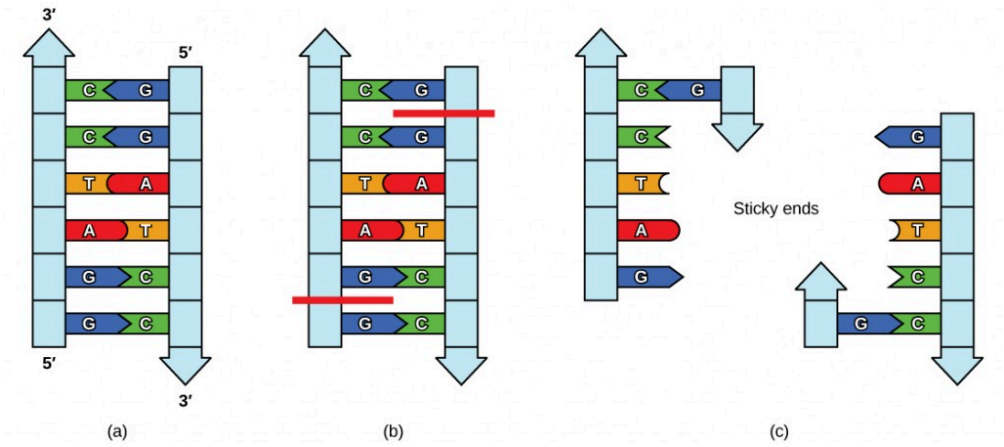


Figure 10.5 In this (a) six-nucleotide restriction enzyme recognition site, notice that the sequence of six nucleotides reads the same in the 5′ to 3′ direction on one strand as it does in the 5′ to 3′ direction on the complementary strand. This is known as a palindrome. (b) The restriction enzyme makes breaks in the DNA strands, and (c) the cut in the DNA results in "sticky ends". Another piece of DNA cut on either end by the same restriction enzyme could attach to these sticky ends and be inserted into the gap made by this cut.

Because these overhangs are capable of coming back together by hydrogen bonding with complementary overhangs on a piece of DNA cut with the same restriction enzyme, these are called "sticky ends." The process of forming hydrogen bonds between complementary sequences on single strands to form double-stranded DNA is called annealing. Addition of an enzyme called DNA ligase, which takes part in DNA replication in cells, permanently joins the DNA fragments when the sticky ends come together. In this way, any DNA fragment can be spliced between the two ends of a plasmid DNA that has been cut with the same restriction enzyme (Figure 10.6).

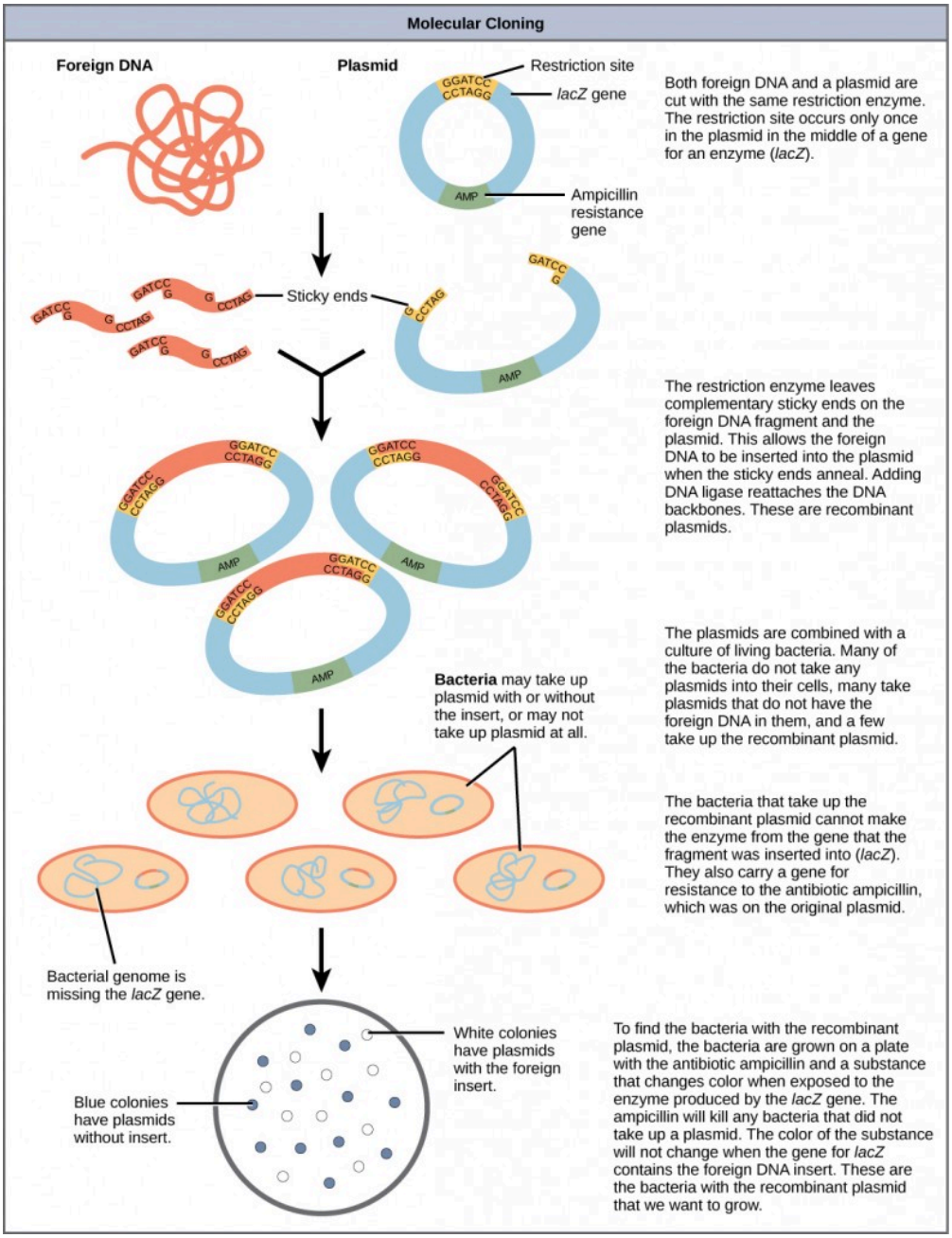


Figure 10.6 This diagram shows the steps involved in molecular cloning.

Plasmids with foreign DNA inserted into them are called recombinant DNA molecules because they contain new combinations of genetic material. Proteins that are produced from recombinant DNA molecules are called recombinant proteins. Not all recombinant plasmids are capable of expressing genes. Plasmids may also be engineered to express proteins only when stimulated by certain environmental factors, so that scientists can control the expression of the recombinant proteins.

Reproductive Cloning

Reproductive cloning is a method used to make a clone or an **identical copy of an entire multicellular organism**. Most multicellular organisms undergo reproduction by sexual means, which involves the

contribution of DNA from two individuals (parents), making it impossible to generate an identical copy or a clone of either parent. Recent advances in biotechnology have made it possible to reproductively clone mammals in the laboratory.

Natural sexual reproduction involves the union, during fertilization, of a sperm and an egg. Each of these gametes is haploid, meaning they contain one set of chromosomes in their nuclei. The resulting cell, or zygote, is then diploid and contains two sets of chromosomes. This cell divides mitotically to produce a multicellular organism. However, the union of just any two cells cannot produce a viable zygote; there are components in the cytoplasm of the egg cell that are essential for the early development of the embryo during its first few cell divisions. Without these provisions, there would be no subsequent development. Therefore, to produce a new individual, both a diploid genetic complement and an egg cytoplasm are required. The approach to producing an artificially cloned individual is to take the egg cell of one individual and to remove the haploid nucleus. Then a diploid nucleus from a body cell of a second individual, the donor, is put into the egg cell. The egg is then stimulated to divide so that development proceeds. This sounds simple, but in fact it takes many attempts before each of the steps is completed successfully.

The first cloned agricultural animal was Dolly, a sheep who was born in 1996. The success rate of reproductive cloning at the time was very low. Dolly lived for six years and died of a lung tumor (Figure 10.7). There was speculation that because the cell DNA that gave rise to Dolly came from an older individual, the age of the DNA may have affected her life expectancy. Since Dolly, several species of animals (such as horses, bulls, and goats) have been successfully cloned.

There have been attempts at producing cloned human embryos as sources of embryonic stem cells. In the procedure, the DNA from an adult human is introduced into a human egg cell, which is then stimulated to divide. The technology is similar to the technology that was used to produce Dolly, but the embryo is never implanted into a surrogate mother. The cells produced are called embryonic stem cells because they have the capacity to develop into many different kinds of cells, such as muscle or nerve cells. The stem cells could be used to research and ultimately provide therapeutic applications, such as replacing damaged tissues. The benefit of cloning in this instance is that the cells used to regenerate new tissues would be a perfect match to the donor of the original DNA. For example, a leukemia patient would not require a sibling with a tissue match for a bone-marrow transplant.

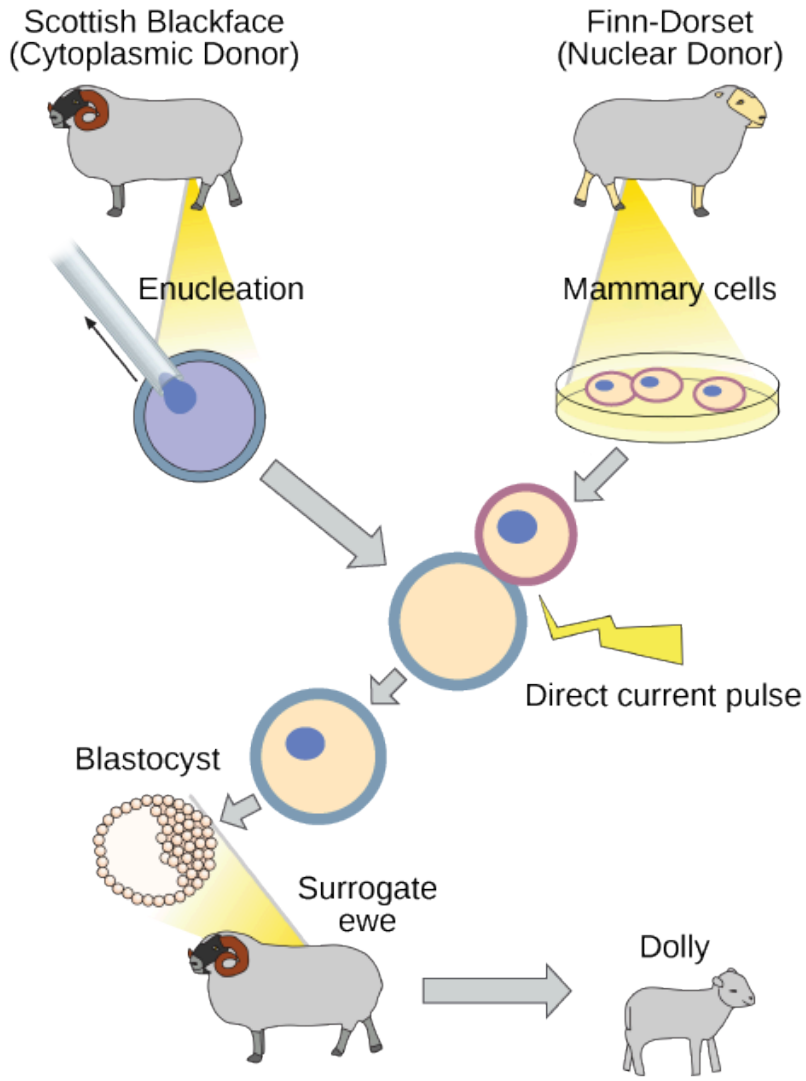


Figure 10.7 Dolly the sheep was the first agricultural animal to be cloned. To create Dolly, the nucleus was removed from a donor egg cell. The enucleated egg was placed next to the other cell, then they were shocked to fuse. They were shocked again to start division. The cells were allowed to divide for several days until an early embryonic stage was reached, before being implanted in a surrogate mother.

Why was Dolly a Finn-Dorset and not a Scottish Blackface sheep?

Because even though the original cell came from a Scottish Blackface sheep and the surrogate mother was a Scottish Blackface, the DNA came from a Finn-Dorset.

Genetic Engineering

Using recombinant DNA technology to modify an organism's DNA to achieve desirable traits is called genetic engineering. Addition of foreign DNA in the form of recombinant DNA vectors that are generated by molecular cloning is the most common method of genetic engineering. An organism that receives the recombinant DNA is called a genetically modified organism (GMO). If the foreign DNA

that is introduced comes from a different species, the host organism is called transgenic. Bacteria, plants, and animals have been genetically modified since the early 1970s for academic, medical, agricultural, and industrial purposes. These applications will be examined in more detail in the next module.

Concept in Action

Watch this short video explaining how scientists create a transgenic animal.

Although the classic methods of studying the function of genes began with a given phenotype and determined the genetic basis of that phenotype, modern techniques allow researchers to start at the DNA sequence level and ask: “What does this gene or DNA element do?” This technique, called reverse genetics, has resulted in reversing the classical genetic methodology. One example of this method is analogous to damaging a body part to determine its function. An insect that loses a wing cannot fly, which means that the wing’s function is flight. The classic genetic method compares insects that cannot fly with insects that can fly, and observes that the non-flying insects have lost wings. Similarly in a reverse genetics approach, mutating or deleting genes provides researchers with clues about gene function. Alternately, reverse genetics can be used to cause a gene to overexpress itself to determine what phenotypic effects may occur.

Section Summary

Nucleic acids can be isolated from cells for the purposes of further analysis by breaking open the cells and enzymatically destroying all other major macromolecules. Fragmented or whole chromosomes can be separated on the basis of size by gel electrophoresis. Short stretches of DNA can be amplified by PCR. DNA can be cut (and subsequently re-spliced together) using restriction enzymes. The molecular and cellular techniques of biotechnology allow researchers to genetically engineer organisms, modifying them to achieve desirable traits.

Cloning may involve cloning small DNA fragments (molecular cloning), or cloning entire organisms (reproductive cloning). In molecular cloning with bacteria, a desired DNA fragment is inserted into a bacterial plasmid using restriction enzymes and the plasmid is taken up by a bacterium, which will then express the foreign DNA. Using other techniques, foreign genes can be inserted into eukaryotic organisms. In each case, the organisms are called transgenic organisms. In reproductive cloning, a donor nucleus is put into an enucleated egg cell, which is then stimulated to divide and develop into an organism.

In reverse genetics methods, a gene is mutated or removed in some way to identify its effect on the phenotype of the whole organism as a way to determine its function.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4615#h5p-62

Glossary

anneal: in molecular biology, the process by which two single strands of DNA hydrogen bond at complementary nucleotides to form a double-stranded molecule

biotechnology: the use of artificial methods to modify the genetic material of living organisms or cells to produce novel compounds or to perform new functions

cloning: the production of an exact copy—specifically, an exact genetic copy—of a gene, cell, or organism

gel electrophoresis: a technique used to separate molecules on the basis of their ability to migrate through a semisolid gel in response to an electric current

genetic engineering: alteration of the genetic makeup of an organism using the molecular methods of biotechnology
genetically modified organism (GMO): an organism whose genome has been artificially changed

plasmid: a small circular molecule of DNA found in bacteria that replicates independently of the main bacterial chromosome; plasmids code for some important traits for bacteria and can be used as vectors to transport DNA into bacteria in genetic engineering applications

polymerase chain reaction (PCR): a technique used to make multiple copies of DNA

recombinant DNA: a combination of DNA fragments generated by molecular cloning that does not exist in nature
strong>recombinant protein: a protein that is expressed from recombinant DNA molecules

restriction enzyme: an enzyme that recognizes a specific nucleotide sequence in DNA and cuts the DNA double strand at that recognition site, often with a staggered cut leaving short single strands or “sticky” ends

reverse genetics: a form of genetic analysis that manipulates DNA to disrupt or affect the product of a gene to analyze the gene’s function

reproductive cloning: cloning of entire organisms

transgenic: describing an organism that receives DNA from a different species

10.2 Biotechnology in Medicine and Agriculture

Learning Objectives

By the end of this section, you will be able to:

- Describe uses of biotechnology in medicine
- Describe uses of biotechnology in agriculture

It is easy to see how biotechnology can be used for medicinal purposes. Knowledge of the genetic makeup of our species, the genetic basis of heritable diseases, and the invention of technology to manipulate and fix mutant genes provides methods to treat diseases. Biotechnology in agriculture can enhance resistance to disease, pests, and environmental stress to improve both crop yield and quality.

Genetic Diagnosis and Gene Therapy

The process of testing for suspected genetic defects before administering treatment is called genetic diagnosis by genetic testing. In some cases in which a genetic disease is present in an individual's family, family members may be advised to undergo genetic testing. For example, mutations in the ***BRCA* genes** may increase the likelihood of developing breast and ovarian cancers in women and some other cancers in women and men. A woman with breast cancer can be screened for these mutations. If one of the high-risk mutations is found, her female relatives may also wish to be screened for that particular mutation, or simply be more vigilant for the occurrence of cancers. Genetic testing is also offered for fetuses (or embryos with in vitro fertilization) to determine the presence or absence of disease-causing genes in families with specific debilitating diseases.

Concept in Action

See how human DNA is extracted for uses such as genetic testing.

Gene therapy is a genetic engineering technique that may one day be used to cure certain genetic

diseases. In its simplest form, it involves **the introduction of a non-mutated gene at a random location in the genome to cure a disease by replacing a protein that may be absent in these individuals because of a genetic mutation**. The non-mutated gene is usually introduced into diseased cells as part of a vector transmitted by a virus, such as an adenovirus, that can infect the host cell and deliver the foreign DNA into the genome of the targeted cell (Figure 10.8). To date, gene therapies have been primarily experimental procedures in humans. A few of these experimental treatments have been successful, but the methods may be important in the future as the factors limiting its success are resolved.

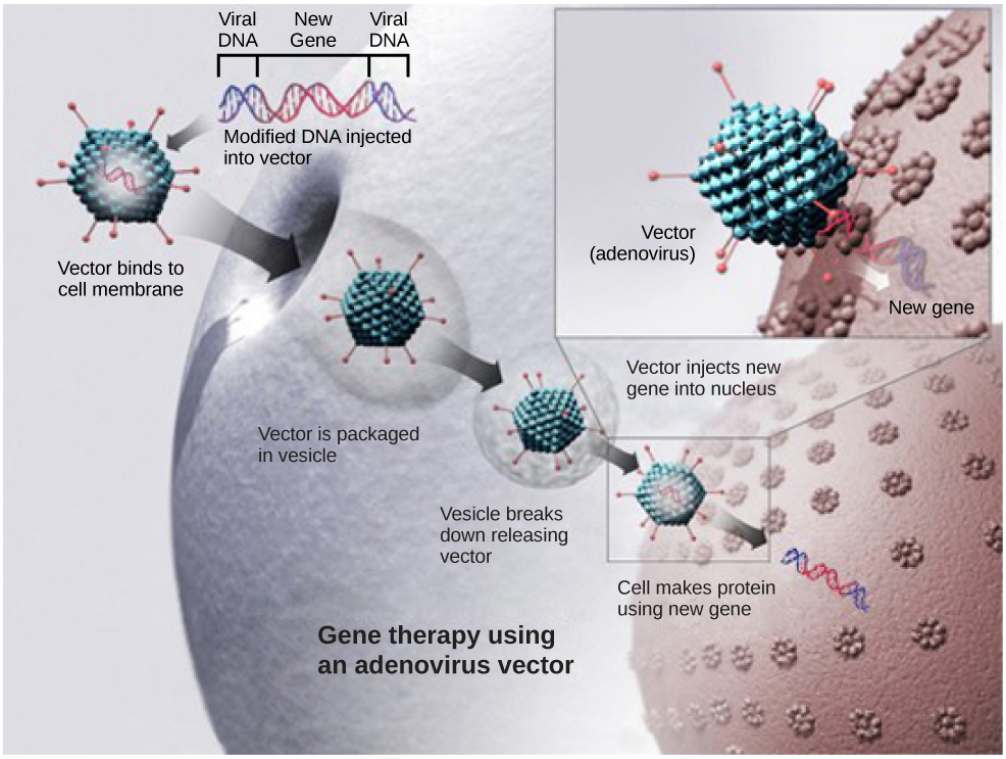


Figure 10.8 This diagram shows the steps involved in curing disease with gene therapy using an adenovirus vector. (credit: modification of work by NIH)

Production of Vaccines, Antibiotics, and Hormones

Traditional vaccination strategies use weakened or inactive forms of microorganisms or viruses to stimulate the immune system. Modern techniques use specific genes of microorganisms cloned into vectors and mass-produced in bacteria to make large quantities of specific substances to stimulate the immune system. The substance is then used as a vaccine. In some cases, such as the H1N1 flu vaccine, genes cloned from the virus have been used to combat the constantly changing strains of this virus.

Antibiotics kill bacteria and are naturally produced by microorganisms such as fungi; penicillin is perhaps the most well-known example. Antibiotics are produced on a large scale by cultivating and manipulating fungal cells. The fungal cells have typically been genetically modified to improve the yields of the antibiotic compound.

Recombinant DNA technology was used to produce large-scale quantities of the human hormone insulin in *E. coli* as early as 1978. Previously, it was only possible to treat diabetes with pig insulin,

which caused allergic reactions in many humans because of differences in the insulin molecule. In addition, human growth hormone (HGH) is used to treat growth disorders in children. The HGH gene was cloned from a cDNA (complementary DNA) library and inserted into *E. coli* cells by cloning it into a bacterial vector.

Transgenic Animals

Although several recombinant proteins used in medicine are successfully produced in bacteria, some proteins need a eukaryotic animal host for proper processing. For this reason, genes have been cloned and expressed in animals such as sheep, goats, chickens, and mice. Animals that have been modified to express recombinant DNA are called transgenic animals (Figure 10.9).

Figure 10.9 It can be seen that two of these mice are transgenic because they have a gene that causes them to fluoresce under a UV light. The non-transgenic mouse does not have the gene that causes fluorescence. (credit: Ingrid Moen et al.)

Several human proteins are expressed in the milk of transgenic sheep and goats. In one commercial example, the FDA has approved a blood anticoagulant protein that is produced in the milk of transgenic goats for use in humans. Mice have been used extensively for expressing and studying the effects of recombinant genes and mutations.

Transgenic Plants

Manipulating the DNA of plants **(creating genetically modified organisms, or GMOs)** has helped to create desirable traits such as disease resistance, herbicide, and pest resistance, better nutritional value, and better shelf life (Figure 10.10). Plants are the most important source of food for the human

population. Farmers developed ways to select for plant varieties with desirable traits long before modern-day biotechnology practices were established.

Figure 10.10 Corn, a major agricultural crop used to create products for a variety of industries, is often modified through plant biotechnology. (credit: Keith Weller, USDA)

Transgenic plants have received DNA from other species. Because they contain unique combinations of genes and are not restricted to the laboratory, transgenic plants and other GMOs are closely monitored by government agencies to ensure that they are fit for human consumption and do not endanger other plant and animal life. Because foreign genes can spread to other species in the environment, particularly in the pollen and seeds of plants, extensive testing is required to ensure ecological stability. Staples like corn, potatoes, and tomatoes were the first crop plants to be genetically engineered.

Transformation of Plants Using *Agrobacterium tumefaciens*

In plants, tumors caused by the bacterium *Agrobacterium tumefaciens* occur by transfer of DNA from the bacterium to the plant. The artificial introduction of DNA into plant cells is more challenging than in animal cells because of the thick plant cell wall. Researchers used the natural transfer of DNA from *Agrobacterium* to a plant host to introduce DNA fragments of their choice into plant hosts. In nature, the disease-causing *A. tumefaciens* have a set of plasmids that contain genes that integrate into the infected plant cell's genome. Researchers manipulate the plasmids to carry the desired DNA fragment and insert it into the plant genome.

The Organic Insecticide *Bacillus thuringiensis*

Bacillus thuringiensis (Bt) is a bacterium that produces protein crystals that are toxic to many insect species that feed on plants. Insects that have eaten Bt toxin stop feeding on the plants within a few hours. After the toxin is activated in the intestines of the insects, death occurs within a couple of days. The crystal toxin genes have been cloned from the bacterium and introduced into plants, therefore allowing plants to produce their own crystal Bt toxin that acts against insects. Bt toxin is safe for the environment and non-toxic to mammals (including humans). As a result, it has been approved for use by organic farmers as a natural insecticide. There is some concern, however, that insects may evolve resistance to the Bt toxin in the same way that bacteria evolve resistance to antibiotics.

FlavrSavr Tomato

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4621#h5p-63

The first GM crop to be introduced into the market was the FlavrSavr Tomato produced in 1994. Molecular genetic technology was used to slow down the process of softening and rotting caused by fungal infections, which led to increased shelf life of the GM tomatoes. Additional genetic modification improved the flavor of this tomato. The FlavrSavr tomato did not successfully stay in the market because of problems maintaining and shipping the crop.

Section Summary

Genetic testing is performed to identify disease-causing genes, and can be used to benefit affected individuals and their relatives who have not developed disease symptoms yet. Gene therapy—by which functioning genes are incorporated into the genomes of individuals with a non-functioning mutant gene—has the potential to cure heritable diseases. Transgenic organisms possess DNA from a different species, usually generated by molecular cloning techniques. Vaccines, antibiotics, and hormones are

examples of products obtained by recombinant DNA technology. Transgenic animals have been created for experimental purposes and some are used to produce some human proteins.

Genes are inserted into plants, using plasmids in the bacterium *Agrobacterium tumefaciens*, which infects plants. Transgenic plants have been created to improve the characteristics of crop plants—for example, by giving them insect resistance by inserting a gene for a bacterial toxin.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4621#h5p-64

Glossary

gene therapy: the technique used to cure heritable diseases by replacing mutant genes with good genes

genetic testing: identifying gene variants in an individual that may lead to a genetic disease in that individual

10.3 Genomics and Proteomics

Learning Objectives

By the end of this section, you will be able to:

- Define genomics and proteomics
- Define whole genome sequencing
- Explain different applications of genomics and proteomics

The study of nucleic acids began with the discovery of DNA, progressed to the study of genes and small fragments, and has now exploded to the field of genomics. Genomics is the study of entire genomes, including the complete set of genes, their nucleotide sequence and organization, and their interactions within a species and with other species. The advances in genomics have been made possible by DNA sequencing technology. Just as information technology has led to Google Maps that enable us to get detailed information about locations around the globe, genomic information is used to create similar maps of the DNA of different organisms.

Mapping Genomes

Genome mapping is the process of finding the location of genes on each chromosome. The maps that are created are comparable to the maps that we use to navigate streets. A genetic map is an illustration that lists genes and their location on a chromosome. Genetic maps provide the big picture (similar to a map of interstate highways) and use genetic markers (similar to landmarks). A genetic marker is a gene or sequence on a chromosome that shows genetic linkage with a trait of interest. The genetic marker tends to be inherited with the gene of interest, and one measure of distance between them is the recombination frequency during meiosis. Early geneticists called this linkage analysis.

Physical maps get into the intimate details of smaller regions of the chromosomes (similar to a detailed road map) (Figure 10.11). A physical map is a representation of the physical distance, in nucleotides, between genes or genetic markers. Both genetic linkage maps and physical maps are required to build a complete picture of the genome. Having a complete map of the genome makes it easier for researchers to study individual genes. Human genome maps help researchers in their efforts to identify human disease-causing genes related to illnesses such as cancer, heart disease, and cystic fibrosis, to name a few. In addition, genome mapping can be used to help identify organisms with beneficial traits, such as microbes with the ability to clean up pollutants or even prevent pollution. Research involving plant genome mapping may lead to methods that produce higher crop yields or to the development of plants that adapt better to climate change.

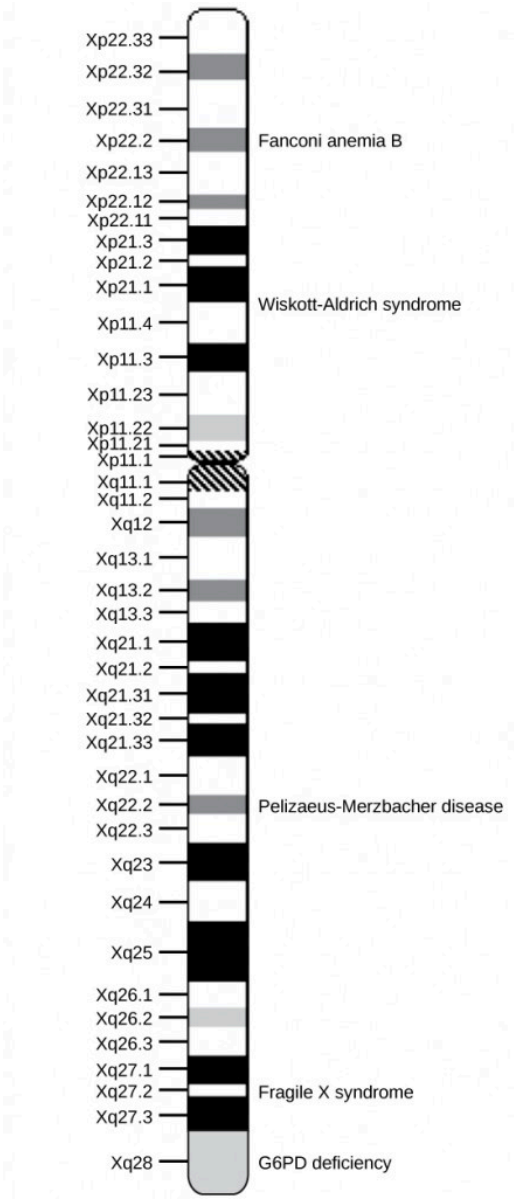


Figure 10.11 This is a physical map of the human X chromosome. (credit: modification of work by NCBI, NIH)

Genetic maps provide the outline, and physical maps provide the details. It is easy to understand why both types of genome-mapping techniques are important to show the big picture. Information obtained from each technique is used in combination to study the genome. Genomic mapping is used with different model organisms that are used for research. Genome mapping is still an ongoing process, and as more advanced techniques are developed, more advances are expected. Genome mapping is similar to completing a complicated puzzle using every piece of available data. Mapping information generated in laboratories all over the world is entered into central databases, such as the National Center for Biotechnology Information (NCBI). Efforts are made to make the information more easily accessible to researchers and the general public. Just as we use global positioning systems instead of paper maps to

navigate through roadways, NCBI allows us to use a genome viewer tool to simplify the data mining process.

Concept in Action

Online Mendelian Inheritance in Man (OMIM) is a searchable online catalog of human genes and genetic disorders. This website shows genome mapping, and also details the history and research of each trait and disorder. Click the link to search for traits (such as handedness) and genetic disorders (such as diabetes).

Whole Genome Sequencing

Although there have been significant advances in the medical sciences in recent years, doctors are still confounded by many diseases and researchers are using whole genome sequencing to get to the bottom of the problem. Whole genome sequencing is a process that determines the DNA sequence of an entire genome. Whole genome sequencing is a brute-force approach to problem solving when there is a genetic basis at the core of a disease. Several laboratories now provide services to sequence, analyze, and interpret entire genomes.

In 2010, whole genome sequencing was used to save a young boy whose intestines had multiple mysterious abscesses. The child had several colon operations with no relief. Finally, a whole genome sequence revealed a defect in a pathway that controls apoptosis (programmed cell death). A bone marrow transplant was used to overcome this genetic disorder, leading to a cure for the boy. He was the first person to be successfully diagnosed using whole genome sequencing.

The first genomes to be sequenced, such as those belonging to viruses, bacteria, and yeast, were smaller in terms of the number of nucleotides than the genomes of multicellular organisms. The genomes of other model organisms, such as the mouse (*Mus musculus*), the fruit fly (*Drosophila melanogaster*), and the nematode (*Caenorhabditis elegans*) are now known. A great deal of basic research is performed in model organisms because the information can be applied to other organisms. A model organism is a species that is studied as a model to understand the biological processes in other species that can be represented by the model organism. For example, fruit flies are able to metabolize alcohol like humans, so the genes affecting sensitivity to alcohol have been studied in fruit flies in an effort to understand the variation in sensitivity to alcohol in humans. Having entire genomes sequenced helps with the research efforts in these model organisms (Figure 10.12).

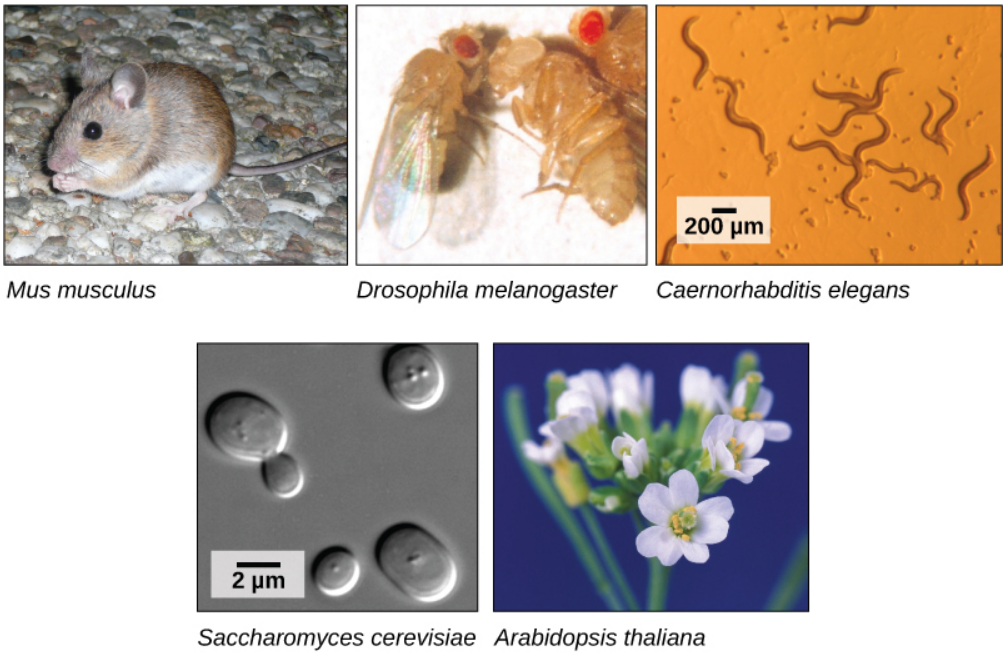


Figure 10.12 Much basic research is done with model organisms, such as the mouse, Mus musculus; the fruit fly, Drosophila melanogaster; the nematode Caenorhabditis elegans; the yeast Saccharomyces cerevisiae; and the common weed, Arabidopsis thaliana. (credit "mouse": modification of work by Florean Fortescue; credit "nematodes": modification of work by "snickclunk"/Flickr; credit "common weed": modification of work by Peggy Greb, USDA; scale-bar data from Matt Russell)

The first human genome sequence was published in 2003. The number of whole genomes that have been sequenced steadily increases and now includes hundreds of species and thousands of individual human genomes.

Applying Genomics

An interactive H5P element has been excluded from this version of the text. You can view it online here: https://opentextbc.ca/biology/?p=4631#h5p-65

The introduction of DNA sequencing and whole genome sequencing projects, particularly the Human Genome Project, has expanded the applicability of DNA sequence information. Genomics is now being used in a wide variety of fields, such as metagenomics, pharmacogenomics, and mitochondrial genomics. The most commonly known application of genomics is to understand and find cures for diseases.

Predicting Disease Risk at the Individual Level

Predicting the risk of disease involves screening and identifying currently healthy individuals by genome analysis at the individual level. Intervention with lifestyle changes and drugs can be recommended before disease onset. However, this approach is most applicable when the problem arises from a single gene mutation. Such defects only account for about 5 percent of diseases found in developed countries. Most of the common diseases, such as heart disease, are multifactorial or polygenic, which refers to a phenotypic characteristic that is determined by two or more genes, and also environmental factors such as diet. In April 2010, scientists at Stanford University published the genome analysis of a healthy individual (Stephen Quake, a scientist at Stanford University, who had his genome sequenced); the analysis predicted his propensity to acquire various diseases. A risk assessment was done to analyze Quake's percentage of risk for 55 different medical conditions. A rare genetic mutation was found that showed him to be at risk for sudden heart attack. He was also predicted to have a 23 percent risk of developing prostate cancer and a 1.4 percent risk of developing Alzheimer's disease. The scientists used databases and several publications to analyze the genomic data. Even though genomic sequencing is becoming more affordable and analytical tools are becoming more reliable, ethical issues surrounding genomic analysis at a population level remain to be addressed. For example, could such data be legitimately used to charge more or less for insurance or to affect credit ratings?

Genome-wide Association Studies

Since 2005, it has been possible to conduct a type of study called a genome-wide association study, or GWAS. A GWAS is a method that identifies differences between individuals in single nucleotide polymorphisms (SNPs) that may be involved in causing diseases. The method is particularly suited to diseases that may be affected by one or many genetic changes throughout the genome. It is very difficult to identify the genes involved in such a disease using family history information. The GWAS method relies on a genetic database that has been in development since 2002 called the International HapMap Project. The HapMap Project sequenced the genomes of several hundred individuals from around the world and identified groups of SNPs. The groups include SNPs that are located near to each other on chromosomes so they tend to stay together through recombination. The fact that the group stays together means that identifying one marker SNP is all that is needed to identify all the SNPs in the group. There are several million SNPs identified, but identifying them in other individuals who have not had their complete genome sequenced is much easier because only the marker SNPs need to be identified.

In a common design for a GWAS, two groups of individuals are chosen; one group has the disease, and the other group does not. The individuals in each group are matched in other characteristics to reduce the effect of confounding variables causing differences between the two groups. For example, the genotypes may differ because the two groups are mostly taken from different parts of the world. Once the individuals are chosen, and typically their numbers are a thousand or more for the study to work, samples of their DNA are obtained. The DNA is analyzed using automated systems to identify large differences in the percentage of particular SNPs between the two groups. Often the study examines a million or more SNPs in the DNA. The results of GWAS can be used in two ways: the genetic differences may be used as markers for susceptibility to the disease in undiagnosed individuals, and the particular genes identified can be targets for research into the molecular pathway of the disease and

potential therapies. An offshoot of the discovery of gene associations with disease has been the formation of companies that provide so-called “personal genomics” that will identify risk levels for various diseases based on an individual’s SNP complement. The science behind these services is controversial.

Because GWAS looks for associations between genes and disease, these studies provide data for other research into causes, rather than answering specific questions themselves. An association between a gene difference and a disease does not necessarily mean there is a cause-and-effect relationship. However, some studies have provided useful information about the genetic causes of diseases. For example, three different studies in 2005 identified a gene for a protein involved in regulating inflammation in the body that is associated with a disease-causing blindness called age-related macular degeneration. This opened up new possibilities for research into the cause of this disease. A large number of genes have been identified to be associated with Crohn’s disease using GWAS, and some of these have suggested new hypothetical mechanisms for the cause of the disease.

Pharmacogenomics

Pharmacogenomics involves evaluating the effectiveness and safety of drugs on the basis of information from an individual’s genomic sequence. Personal genome sequence information can be used to prescribe medications that will be most effective and least toxic on the basis of the individual patient’s genotype. Studying changes in gene expression could provide information about the gene transcription profile in the presence of the drug, which can be used as an early indicator of the potential for toxic effects. For example, genes involved in cellular growth and controlled cell death, when disturbed, could lead to the growth of cancerous cells. Genome-wide studies can also help to find new genes involved in drug toxicity. The gene signatures may not be completely accurate, but can be tested further before pathologic symptoms arise.

Metagenomics

Traditionally, microbiology has been taught with the view that microorganisms are best studied under pure culture conditions, which involves isolating a single type of cell and culturing it in the laboratory. Because microorganisms can go through several generations in a matter of hours, their gene expression profiles adapt to the new laboratory environment very quickly. On the other hand, many species resist being cultured in isolation. Most microorganisms do not live as isolated entities, but in microbial communities known as biofilms. For all of these reasons, pure culture is not always the best way to study microorganisms. Metagenomics is the study of the collective genomes of multiple species that grow and interact in an environmental niche. Metagenomics can be used to identify new species more rapidly and to analyze the effect of pollutants on the environment (Figure 10.13). Metagenomics techniques can now also be applied to communities of higher eukaryotes, such as fish.

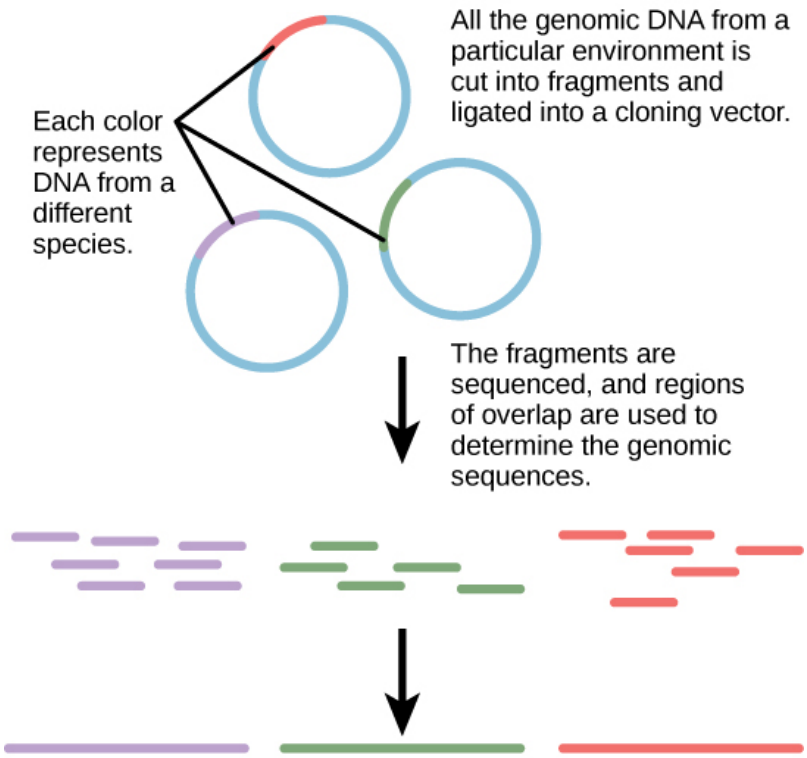


Figure 10.13 Metagenomics involves isolating DNA from multiple species within an environmental niche. The DNA is cut up and sequenced, allowing entire genome sequences of multiple species to be reconstructed from the sequences of overlapping pieces.

Creation of New Biofuels

Knowledge of the genomics of microorganisms is being used to find better ways to harness biofuels from algae and cyanobacteria. The primary sources of fuel today are coal, oil, wood, and other plant products such as ethanol. Although plants are renewable resources, there is still a need to find more alternative renewable sources of energy to meet our population's energy demands. The microbial world is one of the largest resources for genes that encode new enzymes and produce new organic compounds, and it remains largely untapped. This vast genetic resource holds the potential to provide new sources of biofuels (Figure 10.14).

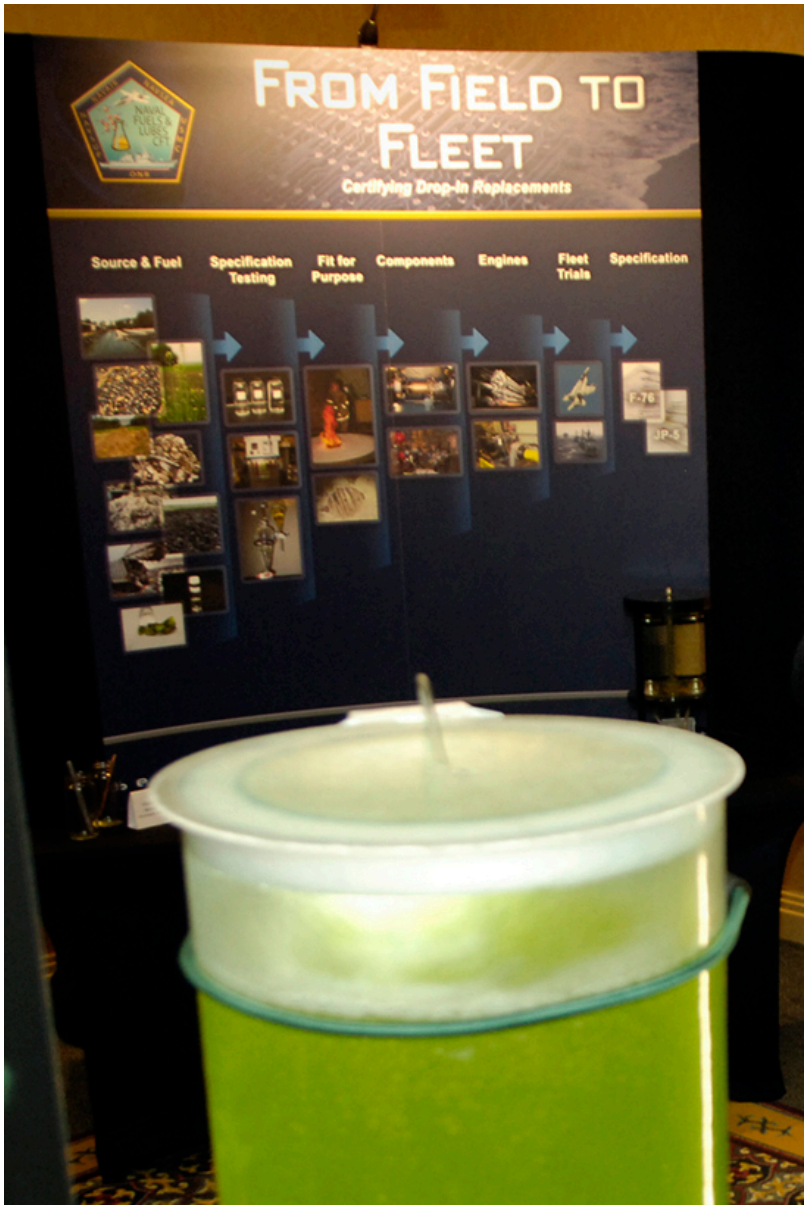


Figure 10.14 Renewable fuels were tested in Navy ships and aircraft at the first Naval Energy Forum. (credit: modification of work by John F. Williams, US Navy)

Mitochondrial Genomics

Mitochondria are intracellular organelles that contain their own DNA. Mitochondrial DNA mutates at a rapid rate and is often used to study evolutionary relationships. Another feature that makes studying the mitochondrial genome interesting is that in most multicellular organisms, the mitochondrial DNA is passed on from the mother during the process of fertilization. For this reason, mitochondrial genomics is often used to trace genealogy.

Genomics in Forensic Analysis

Information and clues obtained from DNA samples found at crime scenes have been used as evidence in court cases, and genetic markers have been used in forensic analysis. Genomic analysis has also become useful in this field. In 2001, the first use of genomics in forensics was published. It was a collaborative effort between academic research institutions and the FBI to solve the mysterious cases of anthrax (Figure 10.15) that was transported by the US Postal Service. Anthrax bacteria were made into an infectious powder and mailed to news media and two U.S. Senators. The powder infected the administrative staff and postal workers who opened or handled the letters. Five people died, and 17 were sickened from the bacteria. Using microbial genomics, researchers determined that a specific strain of anthrax was used in all the mailings; eventually, the source was traced to a scientist at a national biodefense laboratory in Maryland.

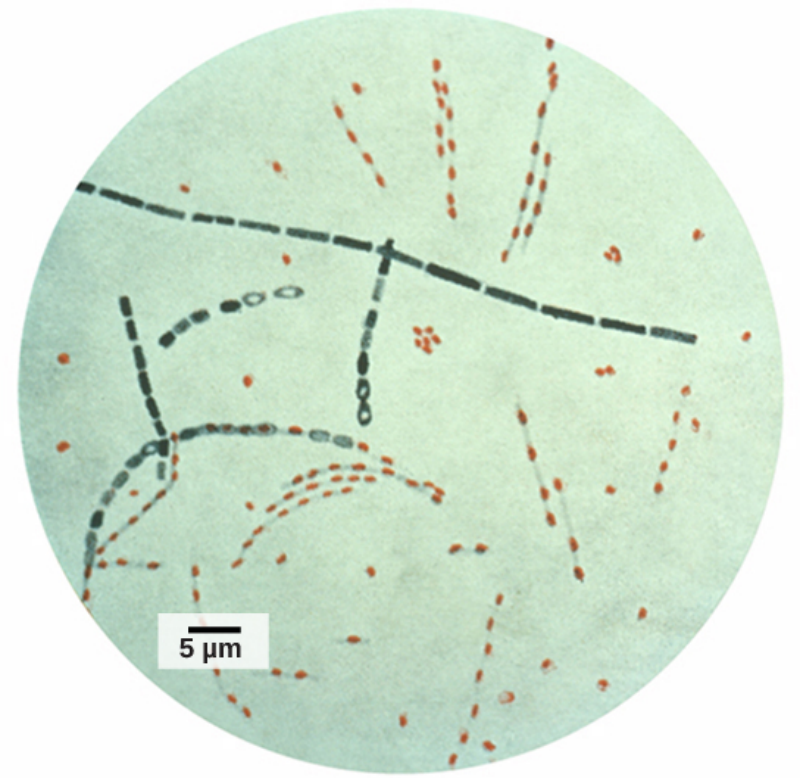


Figure 10.15 Bacillus anthracis is the organism that causes anthrax. (credit: modification of work by CDC; scale-bar data from Matt Russell)

Genomics in Agriculture

Genomics can reduce the trials and failures involved in scientific research to a certain extent, which could improve the quality and quantity of crop yields in agriculture (Figure 10.16). Linking traits to genes or gene signatures helps to improve crop breeding to generate hybrids with the most desirable qualities. Scientists use genomic data to identify desirable traits, and then transfer those traits to a different organism to create a new genetically modified organism, as described in the previous module. Scientists are discovering how genomics can improve the quality and quantity of agricultural

production. For example, scientists could use desirable traits to create a useful product or enhance an existing product, such as making a drought-sensitive crop more tolerant of the dry season.

Figure 10.16 Transgenic agricultural plants can be made to resist disease. These transgenic plums are resistant to the plum pox virus. (credit: Scott Bauer, USDA ARS)

Proteomics

Proteins are the final products of genes that perform the function encoded by the gene. Proteins are composed of amino acids and play important roles in the cell. All enzymes (except ribozymes) are proteins and act as catalysts that affect the rate of reactions. Proteins are also regulatory molecules, and some are hormones. Transport proteins, such as hemoglobin, help transport oxygen to various organs. Antibodies that defend against foreign particles are also proteins. In the diseased state, protein function can be impaired because of changes at the genetic level or because of direct impact on a specific protein.

A proteome is the entire set of proteins produced by a cell type. Proteomes can be studied using the knowledge of genomes because genes code for mRNAs, and the mRNAs encode proteins. The study of the function of proteomes is called proteomics. Proteomics complements genomics and is useful when scientists want to test their hypotheses that were based on genes. Even though all cells in a multicellular organism have the same set of genes, the set of proteins produced in different tissues is different and dependent on gene expression. Thus, the genome is constant, but the proteome varies and is dynamic within an organism. In addition, RNAs can be alternatively spliced (cut and pasted to create novel combinations and novel proteins), and many proteins are modified after translation. Although the genome provides a blueprint, the final architecture depends on several factors that can change the progression of events that generate the proteome.

Genomes and proteomes of patients suffering from specific diseases are being studied to understand the genetic basis of the disease. The most prominent disease being studied with proteomic approaches is cancer (Figure 10.17). Proteomic approaches are being used to improve the screening and early

detection of cancer; this is achieved by identifying proteins whose expression is affected by the disease process. An individual protein is called a biomarker, whereas a set of proteins with altered expression levels is called a protein signature. For a biomarker or protein signature to be useful as a candidate for early screening and detection of a cancer, it must be secreted in body fluids such as sweat, blood, or urine, so that large-scale screenings can be performed in a noninvasive fashion. The current problem with using biomarkers for the early detection of cancer is the high rate of false-negative results. A false-negative result is a negative test result that should have been positive. In other words, many cases of cancer go undetected, which makes biomarkers unreliable. Some examples of protein biomarkers used in cancer detection are CA-125 for ovarian cancer and PSA for prostate cancer. Protein signatures may be more reliable than biomarkers to detect cancer cells. Proteomics is also being used to develop individualized treatment plans, which involves the prediction of whether or not an individual will respond to specific drugs and the side effects that the individual may have. Proteomics is also being used to predict the possibility of disease recurrence.

Figure 10.17 This machine is preparing to do a proteomic pattern analysis to identify specific cancers so that an accurate cancer prognosis can be made. (credit: Dorie Hightower, NCI, NIH)

The National Cancer Institute has developed programs to improve the detection and treatment of cancer. The Clinical Proteomic Technologies for Cancer and the Early Detection Research Network are efforts to identify protein signatures specific to different types of cancers. The Biomedical Proteomics Program is designed to identify protein signatures and design effective therapies for cancer patients.

Section Summary

Genome mapping is similar to solving a big, complicated puzzle with pieces of information coming from laboratories all over the world. Genetic maps provide an outline for the location of genes within a genome, and they estimate the distance between genes and genetic markers on the basis of the recombination frequency during meiosis. Physical maps provide detailed information about the

physical distance between the genes. The most detailed information is available through sequence mapping. Information from all mapping and sequencing sources is combined to study an entire genome.

Whole genome sequencing is the latest available resource to treat genetic diseases. Some doctors are using whole genome sequencing to save lives. Genomics has many industrial applications, including biofuel development, agriculture, pharmaceuticals, and pollution control.

Imagination is the only barrier to the applicability of genomics. Genomics is being applied to most fields of biology; it can be used for personalized medicine, prediction of disease risks at an individual level, the study of drug interactions before the conduction of clinical trials, and the study of microorganisms in the environment as opposed to the laboratory. It is also being applied to the generation of new biofuels, genealogical assessment using mitochondria, advances in forensic science, and improvements in agriculture.

Proteomics is the study of the entire set of proteins expressed by a given type of cell under certain environmental conditions. In a multicellular organism, different cell types will have different proteomes, and these will vary with changes in the environment. Unlike a genome, a proteome is dynamic and under constant flux, which makes it more complicated and more useful than the knowledge of genomes alone.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4631#h5p-66

Glossary

biomarker: an individual protein that is uniquely produced in a diseased state

genetic map: an outline of genes and their location on a chromosome that is based on recombination frequencies between markers

genomics: the study of entire genomes, including the complete set of genes, their nucleotide sequence and organization, and their interactions within a species and with other species

metagenomics: the study of the collective genomes of multiple species that grow and interact in an environmental niche

model organism: a species that is studied and used as a model to understand the biological processes in other species represented by the model organism

pharmacogenomics: the study of drug interactions with the genome or proteome; also called toxicogenomics

physical map: a representation of the physical distance between genes or genetic markers

protein signature: a set of over- or under-expressed proteins characteristic of cells in a particular diseased tissue

proteomics: study of the function of proteomes

whole genome sequencing: a process that determines the nucleotide sequence of an entire genome

UNIT 4: ANIMAL STRUCTURE AND FUNCTION

Unit 4: Animal Structure and Function includes the following chapters:

Note for instructors, students, and other users: This version of *Concepts of Biology* contain two sets of anatomy and physiology chapters and PowerPoint presentations. Chapters 11–13 and associated PPTs cover materials briefly, while Chapters 14–24 and associated PPTs go into more depth. Let your interests and course objectives guide your use of these materials.

Chapter 11: Introduction to the Body's Systems

Figure 11.1 An arctic fox is a complex animal, well adapted to its environment. (credit: Keith Morehouse, USFWS)

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

Search for Key Points in Chapter 11

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4636#h5p-67

11.1 Homeostasis and Osmoregulation

Learning Objectives

By the end of this section, you will be able to:

- Explain the concept of homeostasis
- Describe thermoregulation of endothermic and ectothermic animals
- Explain how the kidneys serve as the main osmoregulatory organs in the human body

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

Homeostasis

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

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

Thermoregulation

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

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

Concept in Action

Watch this Discovery Channel video on thermoregulation to see illustrations of the process in a variety of animals.

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

Ectotherms and endotherms use their circulatory systems to help maintain body temperature. Vasodilation, the opening up of arteries to the skin by relaxation of their smooth muscles, brings more blood and heat to the body surface, facilitating radiation and evaporative heat loss, cooling the body. Vasoconstriction, the narrowing of blood vessels to the skin by contraction of their smooth muscles, reduces blood flow in peripheral blood vessels, forcing blood toward the core and vital organs,

conserving heat. Some animals have adaptions to their circulatory system that enable them to transfer heat from arteries to veins that are flowing next to each other, warming blood returning to the heart. This is called a countercurrent heat exchange; it prevents the cold venous blood from cooling the heart and other internal organs. The countercurrent adaptation is found in dolphins, sharks, bony fish, bees, and hummingbirds.

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

Thermoregulation is coordinated by the nervous system (Figure 11.2). The processes of temperature control are centered in the hypothalamus of the advanced animal brain. The hypothalamus maintains the set point for body temperature through reflexes that cause vasodilation or vasoconstriction and shivering or sweating. The sympathetic nervous system under control of the hypothalamus directs the responses that effect the changes in temperature loss or gain that return the body to the set point. The set point may be adjusted in some instances. During an infection, compounds called pyrogens are produced and circulate to the hypothalamus resetting the thermostat to a higher value. This allows the body's temperature to increase to a new homeostatic equilibrium point in what is commonly called a fever. The increase in body heat makes the body less optimal for bacterial growth and increases the activities of cells so they are better able to fight the infection.

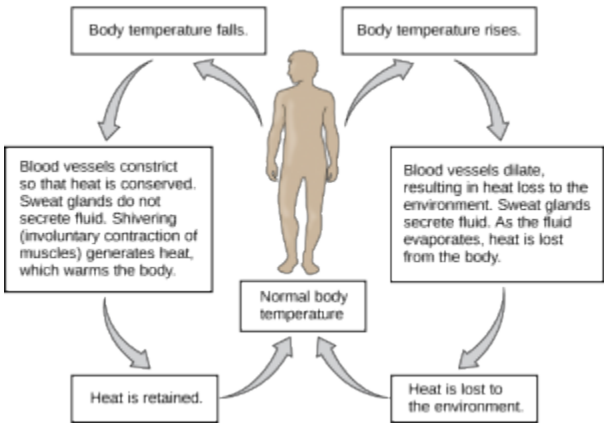


Figure 11.2 The body is able to regulate temperature in response to signals from the nervous system.

When bacteria are destroyed by leukocytes, pyrogens are released into the blood. Pyrogens reset the body's thermostat to a higher temperature, resulting in fever. How might pyrogens cause the body temperature to rise?

<!–Pyrogens increase body temperature by causing the blood vessels to constrict, inducing shivering, and stopping sweat glands from secreting fluid.–>

Osmoregulation

Osmoregulation is the process of maintaining salt and water balance (osmotic balance) across

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

The body does not exist in isolation. There is a constant input of water and electrolytes into the system. Excess water, electrolytes, and wastes are transported to the kidneys and excreted, helping to maintain osmotic balance. Insufficient fluid intake results in fluid conservation by the kidneys. Biological systems constantly interact and exchange water and nutrients with the environment by way of consumption of food and water and through excretion in the form of sweat, urine, and feces. Without a mechanism to regulate osmotic pressure, or when a disease damages this mechanism, there is a tendency to accumulate toxic waste and water, which can have dire consequences.

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

Excretory System

The human excretory system functions to remove waste from the body through the skin as sweat, the lungs in the form of exhaled carbon dioxide, and through the urinary system in the form of urine. All three of these systems participate in osmoregulation and waste removal. Here we focus on the urinary system, which is comprised of the paired kidneys, the ureter, urinary bladder and urethra (Figure 11.3). The kidneys are a pair of bean-shaped structures that are located just below the liver in the body cavity. Each of the kidneys contains more than a million tiny units called nephrons that filter blood containing the metabolic wastes from cells. All the blood in the human body is filtered about 60 times a day by the kidneys. The nephrons remove wastes, concentrate them, and form urine that is collected in the bladder.

Internally, the kidney has three regions—an outer cortex, a medulla in the middle, and the renal pelvis, which is the expanded end of the ureter. The renal cortex contains the nephrons—the functional unit of the kidney. The renal pelvis collects the urine and leads to the ureter on the outside of the kidney. The ureters are urine-bearing tubes that exit the kidney and empty into the urinary bladder.

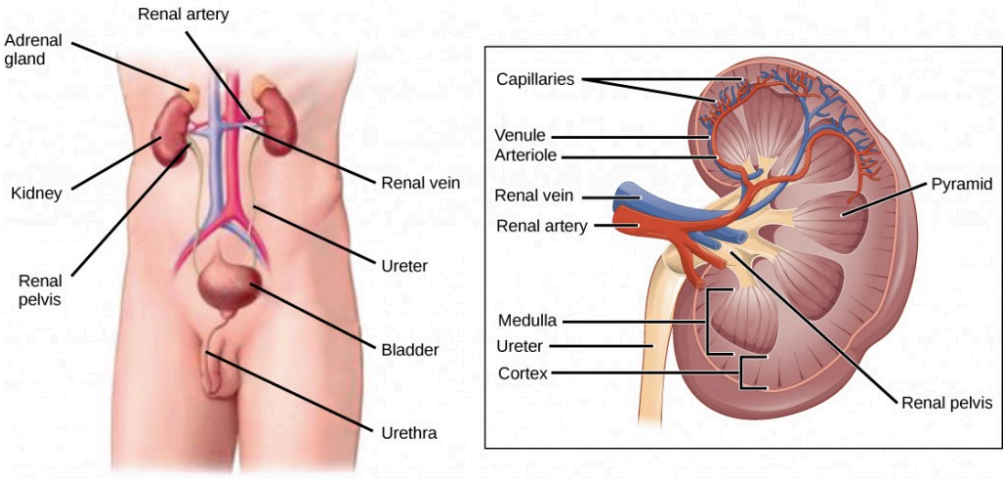


Figure 11.3 The human excretory system is made up of the kidneys, ureter, urinary bladder, and urethra. The kidneys filter blood and form urine, which is stored in the bladder until it is eliminated through the urethra. On the right, the internal structure of the kidney is shown. (credit: modification of work by NCI, NIH)

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

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

Dialysis Technician

Dialysis is a medical process of removing wastes and excess water from the blood by diffusion and ultrafiltration. When kidney function fails, dialysis must be done to artificially rid the body of wastes and fluids. This is a vital process to keep patients alive. In some cases, the patients undergo artificial dialysis until they are eligible for a kidney transplant. In others who are not candidates for kidney transplants, dialysis is a lifelong necessity.

Dialysis technicians typically work in hospitals and clinics. While some roles in this field include equipment development and maintenance, most dialysis technicians work in direct patient care. Their on-the-job duties, which typically occur under the direct supervision of a registered nurse, focus on providing dialysis treatments. This can include reviewing patient history and current condition, assessing and responding to patient needs before and during treatment, and monitoring the dialysis process. Treatment may include taking and reporting a patient's vital signs, preparing solutions and equipment to ensure accurate and sterile procedures.

Section Summary

Homeostasis is a dynamic equilibrium that is maintained in body tissues and organs. It is dynamic because it is constantly adjusting to the changes that the systems encounter. It is an equilibrium because body functions are kept within a normal range, with some fluctuations around a set point. The kidneys are the main osmoregulatory organs in mammalian systems; they function to filter blood and maintain the dissolved ion concentrations of body fluids. They are made up internally of three distinct regions—the cortex, medulla, and pelvis. The blood vessels that transport blood into and out of the kidneys arise from and merge with the aorta and inferior vena cava, respectively. The nephron is the functional unit of the kidney, which actively filters blood and generates urine. The urine leaves the kidney through the ureter and is stored in the urinary bladder. Urine is voided from the body through the urethra.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4641#h5p-68

Glossary

ectotherm: an organism that relies primarily on environmental heat sources to maintain its body temperature
endotherm: an organism that relies primarily on internal heat sources to maintain its body temperature
interstitial fluid: the fluid found between cells in the body, similar in constitution to the fluid component of blood, but without the high concentrations of proteins
kidney: the organ that performs excretory and osmoregulatory functions
nephron: the functional unit of the kidney
osmoregulation: the mechanism by which water and solute concentrations are maintained at desired levels
osmotic balance: the appropriate values of water and solute concentrations for a healthy organism
renal artery: the artery that delivers blood to the kidney

renal vein: the vein that drains blood from the kidney
set point: the target value of a physiological state in homeostasis
ureter: the urine-bearing tubes coming out of the kidney
urethra: the tube that conducts urine from the urinary bladder to the external environment
urinary bladder: the structure that the ureters empty the urine into the appropriate values of water and solute concentrations for a healthy organism

11.2 Digestive System

Learning Objectives

By the end of this section, you will be able to:

- Explain the processes of digestion and absorption
- Explain the specialized functions of the organs involved in processing food in the body
- Describe the ways in which organs work together to digest food and absorb nutrients
- Describe the essential nutrients required for cellular function that cannot be synthesized by the animal body
- Describe how excess carbohydrates and energy are stored in the body

All living organisms need nutrients to survive. While plants can obtain nutrients from their roots and the energy molecules required for cellular function through the process of photosynthesis, animals obtain their nutrients by the consumption of other organisms. At the cellular level, the biological molecules necessary for animal function are amino acids, lipid molecules, nucleotides, and simple sugars. However, the food consumed consists of protein, fat, and complex carbohydrates. Animals must convert these macromolecules into the simple molecules required for maintaining cellular function. The conversion of the food consumed to the nutrients required is a multistep process involving digestion and absorption. During digestion, food particles are broken down to smaller components, which are later absorbed by the body. This happens by both physical means, such as chewing, and by chemical means.

One of the challenges in human nutrition is maintaining a balance between food intake, storage, and energy expenditure. Taking in more food energy than is used in activity leads to storage of the excess in the form of fat deposits. The rise in obesity and the resulting diseases like type 2 diabetes makes understanding the role of diet and nutrition in maintaining good health all the more important.

The Human Digestive System

The process of digestion begins in the mouth with the intake of food. The teeth play an important role in masticating (chewing) or physically breaking food into smaller particles. The enzymes present in saliva also begin to chemically break down food. The food is then swallowed and enters the esophagus—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 toward the stomach. The stomach contents are extremely acidic, with a pH between 1.5 and 2.5. This acidity kills microorganisms, breaks down food tissues, and activates digestive enzymes. Further breakdown of food takes place in the small intestine where bile produced by the liver, and enzymes produced by the

small intestine and the pancreas, continue the process of digestion. The smaller molecules are absorbed into the blood stream through the epithelial cells lining the walls of the small intestine. 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 anus.

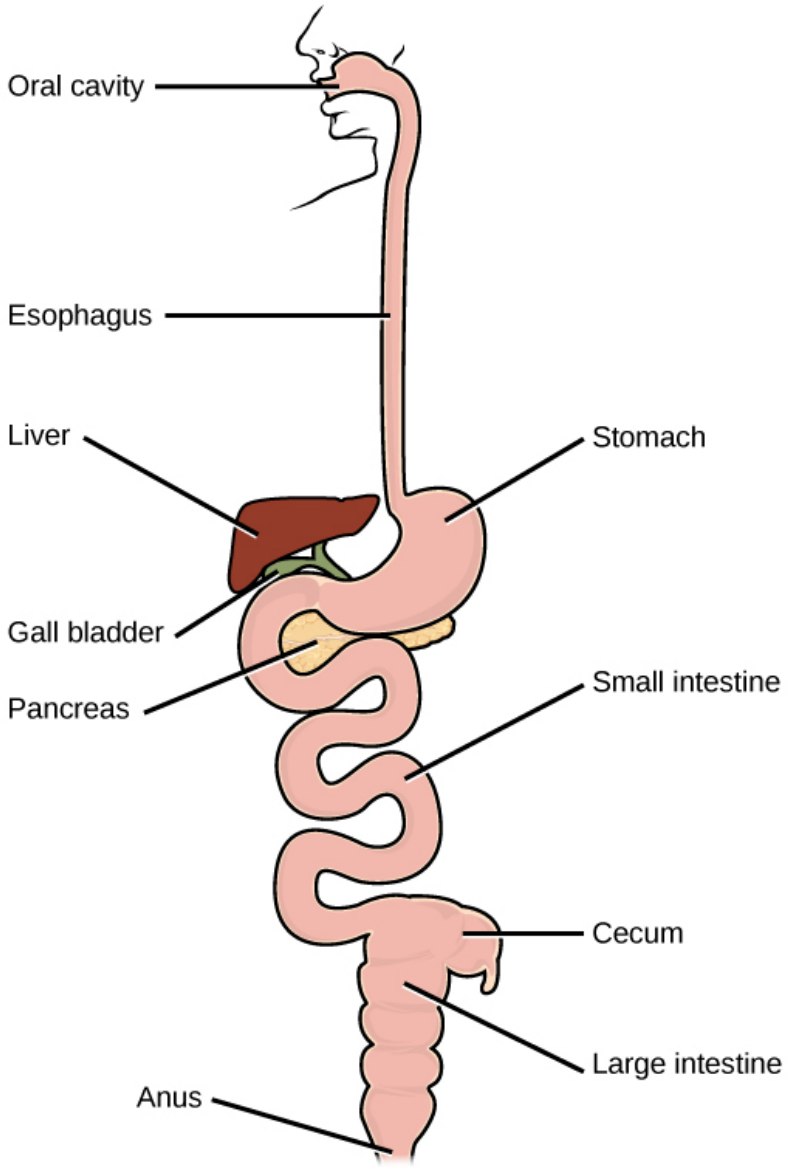


Figure 11.4 The components of the human digestive system are shown.

Oral Cavity

Both physical and chemical digestion begin in the mouth or oral cavity, which is the point of entry of food into the digestive system. The food is broken into smaller particles by mastication, the chewing action of the teeth. All mammals have teeth and can chew their food to begin the process of physically breaking it down into smaller particles.

The chemical process of digestion begins during chewing as food mixes with saliva, produced by the salivary glands (Figure 11.5). Saliva contains mucus that moistens food and buffers the pH of the food. Saliva also contains lysozyme, which has antibacterial action. It also contains an enzyme called salivary amylase that begins the process of converting starches in the food into a disaccharide called maltose. Another enzyme called lipase is produced by cells in the tongue to break down fats. 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 esophagus and the trachea. The esophagus leads to the stomach and the trachea leads to the lungs. The epiglottis is a flap of tissue that covers the tracheal opening during swallowing to prevent food from entering the lungs.

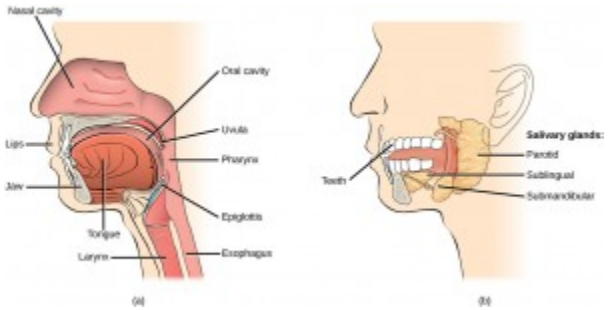


Figure 11.5 (a) Digestion of food begins in the mouth. (b) Food is masticated by teeth and moistened by saliva secreted from the salivary glands. Enzymes in the saliva begin to digest starches and fats. With the help of the tongue, the resulting bolus is moved into the esophagus by swallowing. (credit: modification of work by Mariana Ruiz Villareal)

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 peristalsis that pushes the food toward the stomach. The peristaltic wave is unidirectional—it moves food from the mouth to the stomach, and reverse movement is not possible, except in the case of the vomit reflex. The peristaltic movement of the esophagus is an involuntary reflex; it takes place in response to the act of swallowing.

Ring-like muscles called sphincters form valves in the digestive system. The gastro-esophageal sphincter (or cardiac 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. Acid reflux or “heartburn” occurs when the acidic digestive juices escape into the esophagus.

Stomach

A large part of protein digestion occurs in the stomach (Figure 11.7). The stomach is a saclike organ that secretes gastric digestive juices.

Protein digestion is carried out by an enzyme called pepsin in the stomach chamber. The highly acidic environment kills many microorganisms in the food and, combined with the action of the enzyme pepsin, results in the catabolism of protein in the food. Chemical digestion is facilitated by the churning action of the stomach caused by contraction and relaxation of smooth muscles. The partially digested food and gastric juice mixture is called chyme. 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 hormones, stomach distension and muscular reflexes that influence the pyloric sphincter.

The stomach lining is unaffected by pepsin and the acidity because pepsin is released in an inactive form and the stomach has a thick mucus lining that protects the underlying tissue.

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 top surface of each villus has many microscopic projections called microvilli. The epithelial cells of these structures absorb nutrients from the digested food and release them to the bloodstream on the other side. The villi and microvilli, with their many folds, increase the surface area of the small intestine and increase absorption efficiency of the nutrients.

The human small intestine is over 6 m (19.6 ft) long and is divided into three parts: the duodenum, the jejunum and the ileum. The duodenum is separated from the stomach by the pyloric sphincter. The chyme is mixed with pancreatic juices, an alkaline solution rich in bicarbonate that neutralizes the acidity of chyme from the stomach. Pancreatic juices contain several digestive enzymes that break down starches, disaccharides, proteins, and fats. Bile is produced in the liver and stored and concentrated in the gallbladder; it enters the duodenum through the bile duct. Bile contains bile salts, which make lipids accessible to the water-soluble enzymes. The monosaccharides, amino acids, bile salts, vitamins, and other nutrients are absorbed by the cells of the intestinal lining.

The undigested food is sent to the colon from the ileum via peristaltic movements. 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 has a minor role in immunity.

Large Intestine

The large intestine reabsorbs the water from indigestible food material and processes the waste material (Figure 11.6). The human large intestine is much smaller in length compared to the small intestine but larger in diameter. It has three parts: the cecum, the colon, and the rectum. The cecum joins the ileum to the colon and is the receiving pouch for the waste matter. The colon is home to many bacteria or

“intestinal flora” that aid in the digestive processes. The colon has 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.

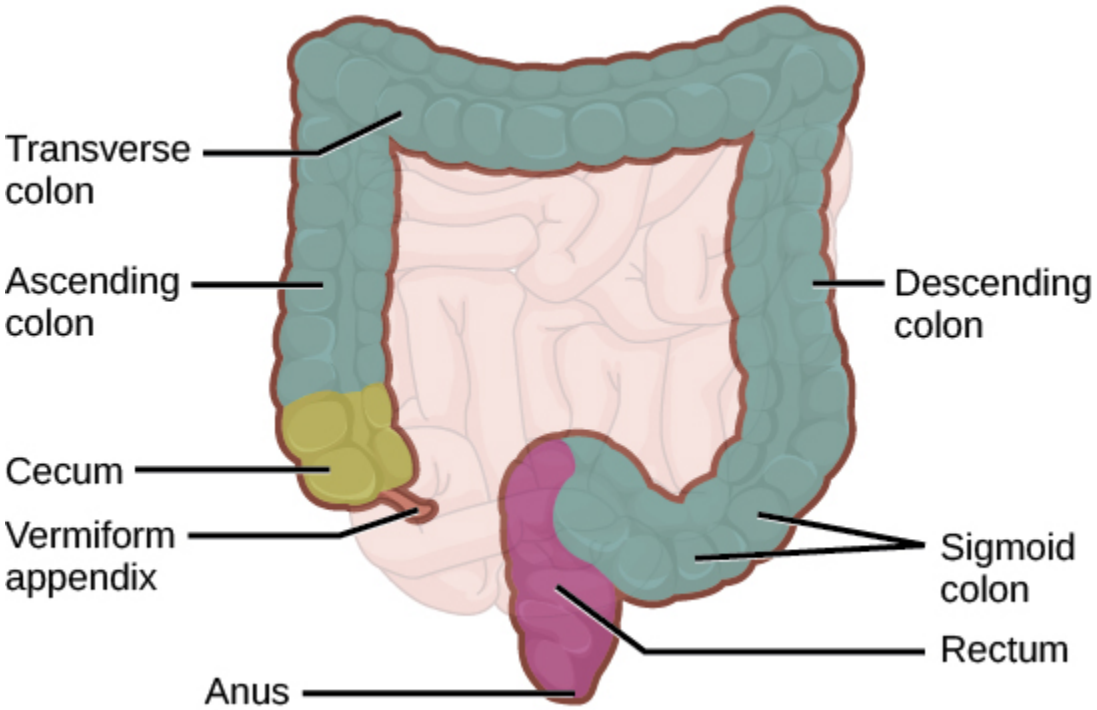


Figure 11.6 The large intestine reabsorbs water from undigested food and stores waste until it is eliminated. (credit: modification of work by Mariana Ruiz Villareal)

The rectum (Figure 11.6) stores 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 regulate the exit of feces, 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 add secretions and enzymes that break down food into nutrients. Accessory organs include the salivary glands, the liver, the pancreas, and the gall bladder. The secretions of the liver, pancreas, and gallbladder are regulated by hormones in response to food consumption.

The liver is the largest internal organ in humans and it plays an important role in digestion of fats and detoxifying blood. The liver produces bile, a digestive juice that is required for the breakdown of fats in the duodenum. The liver also processes the absorbed vitamins and fatty acids and synthesizes many plasma proteins. The gallbladder is a small organ that aids the liver by storing bile and concentrating bile salts.

The pancreas secretes bicarbonate that neutralizes the acidic chyme and a variety of enzymes for the digestion of protein and carbohydrates.

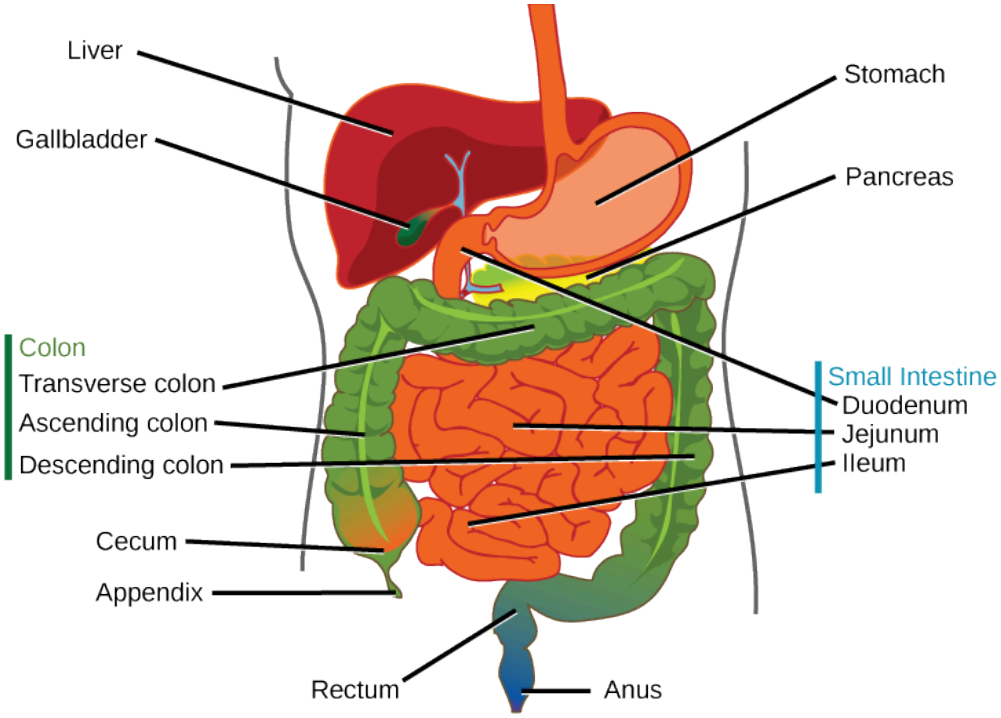


Figure 11.7 The stomach has an extremely acidic environment where most of the protein gets digested. (credit: modification of work by Mariana Ruiz Villareal)

Nutrition

The following video is primarily about water soluble vitamins such as vitamin B and C their roles, especially in energy metabolism. Some of the more common and obscure minerals found in vitamins are also identified.

Vitamin Types

An interactive H5P element has been excluded from this version of the text. You can view it online here: *https://opentextbc.ca/biology/?p=4649#h5p-69*

And the next video is an introduction to another category of vitamins, the fat soluble group such as vitamin E, D and K.

Fat Soluble Vitamins

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4649#h5p-70

The human diet should be well balanced to provide nutrients required for bodily function and the minerals and vitamins required for maintaining structure and regulation necessary for good health and reproductive capability (Figure 11.8).

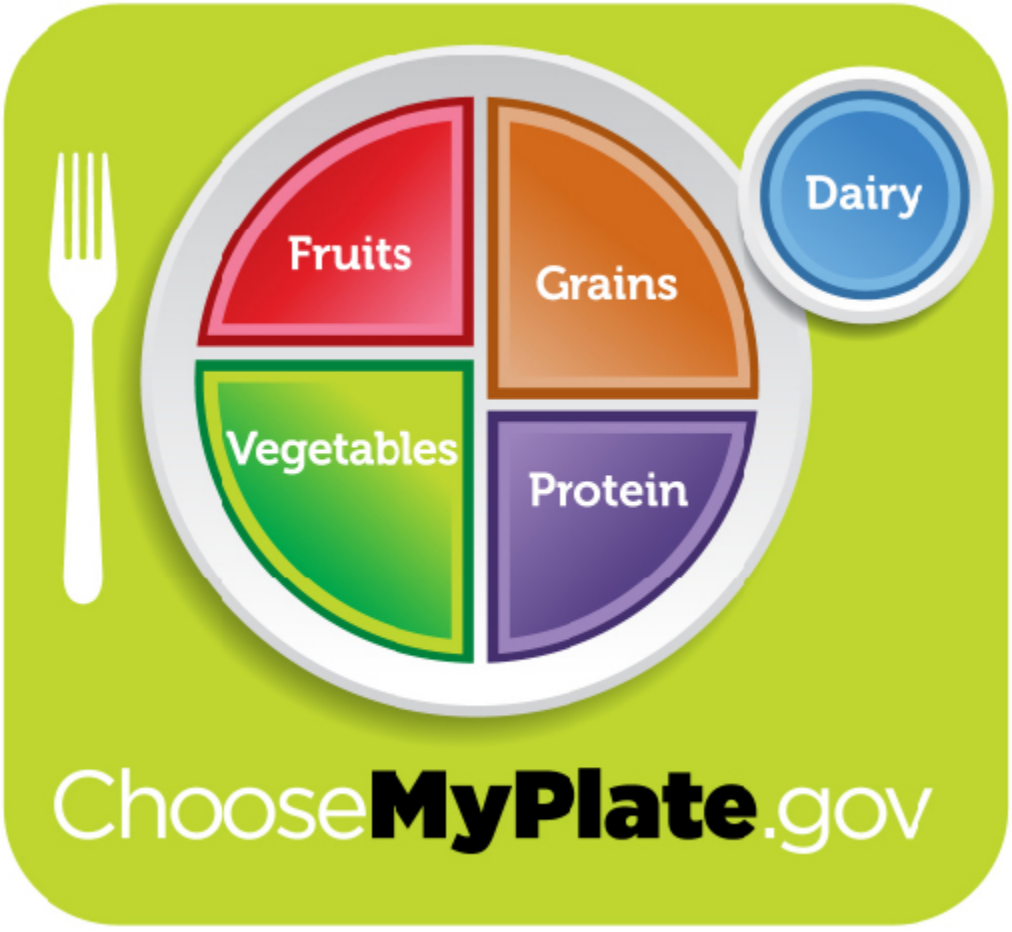


Figure 11.8 For humans, a balanced diet includes fruits, vegetables, grains, protein, and dairy. (credit: USDA)

Concept in Action

Explore this interactive United States Department of Agriculture website to learn more about each food group and the recommended daily amounts.

The organic molecules required for building cellular material and tissues must come from food. During digestion, digestible carbohydrates are ultimately broken down into glucose and used to provide energy within the cells of the body. Complex carbohydrates, including polysaccharides, can be broken down into glucose through biochemical modification; however, humans do not produce the enzyme necessary to digest cellulose (fiber). The intestinal flora in the human gut are able to extract some nutrition from these plant fibers. These plant fibers are known as dietary fiber and are an important component of the diet. The excess sugars in the body are converted into glycogen and stored for later use in the liver and muscle tissue. Glycogen stores are used to fuel prolonged exertions, such as long-distance running, and to provide energy during food shortage. Fats are stored under the skin of mammals for insulation and energy reserves.

Proteins in food are broken down during digestion and the resulting amino acids are absorbed. All of the proteins in the body must be formed from these amino-acid constituents; no proteins are obtained directly from food.

Fats add flavor to food and promote a sense of satiety or fullness. Fatty foods are also significant sources of energy, and fatty acids are required for the construction of lipid membranes. Fats are also required in the diet to aid the absorption of fat-soluble vitamins and the production of fat-soluble hormones.

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

The fatty acids omega-3 alpha-linolenic acid and 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. Many of these assist enzymes in their function and, for this reason, are called coenzymes. Absence or low levels of vitamins can have a dramatic effect on health. Minerals are another set of inorganic essential nutrients that must be obtained from food. Minerals perform many functions, from muscle and nerve function, to acting as enzyme cofactors. 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.

Obesity

With obesity at high rates in the United States, there is a public health focus on reducing obesity and associated health risks, which include diabetes, colon and breast cancer, and cardiovascular disease. How does the food consumed contribute to obesity?

Fatty foods are calorie-dense, meaning that they have more calories per unit mass than carbohydrates or proteins. One gram of carbohydrates has four calories, one gram of protein has four calories, and one gram of fat has nine calories. Animals tend to seek lipid-rich food for their higher energy content. Greater amounts of food energy taken in than the body's requirements will result in storage of the excess in fat deposits.

Excess carbohydrate is used by the liver to synthesize glycogen. When glycogen stores are full, additional glucose is converted into fatty acids. These fatty acids are stored in adipose tissue cells—the fat cells in the mammalian body whose primary role is to store fat for later use.

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, in 2010 First Lady Michelle Obama 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 in 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. Visit www.letsmove.gov to learn more.

Section Summary

There are many organs that work together to digest food and absorb nutrients. The mouth is the point of ingestion and the location where both mechanical and chemical breakdown of food begins. Saliva contains an enzyme called amylase that breaks down carbohydrates. The food bolus travels through the esophagus by peristaltic movements to the stomach. The stomach has an extremely acidic environment. The enzyme pepsin digests protein in the stomach. Further digestion and absorption take place in the small intestine. The large intestine reabsorbs water from the undigested food and stores waste until elimination.

Carbohydrates, proteins, and fats are the primary components of food. Some essential nutrients are required for cellular function but cannot be produced by the animal body. These include vitamins, minerals, some fatty acids, and some amino acids. Food intake in more than necessary amounts is stored as glycogen in the liver and muscle cells, and in adipose tissue. Excess adipose storage can lead to obesity and serious health problems.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4649#h5p-71

Glossary

amylase: an enzyme found in saliva and secreted by the pancreas that converts carbohydrates to maltose
anus: the exit point of the digestive system for waste material
bile: a digestive juice produced by the liver; important for digestion of lipids
bolus: a mass of food resulting from chewing action and wetting by saliva
colon: the largest portion of the large intestine consisting of the ascending colon, transverse colon, and descending colon
chyme: a mixture of partially digested food and stomach juices
esophagus: a tubular organ that connects the mouth to the stomach
essential nutrient: a nutrient that cannot be synthesized by the body; it must be obtained from food
gallbladder: the organ that stores and concentrates bile
large intestine: a digestive system organ that reabsorbs water from undigested material and processes waste matter
liver: an organ that produces bile for digestion and processes vitamins and lipids
mineral: an inorganic, elemental molecule that carries out important roles in the body
oral cavity: the point of entry of food into the digestive system
pancreas: a gland that secretes digestive juices
pepsin: an enzyme found in the stomach whose main role is protein digestion
peristalsis: wave-like movements of muscle tissue
rectum: the area of the body where feces is stored until elimination
salivary gland: one of three pairs of exocrine glands in the mammalian mouth that secretes saliva, a mix of watery mucus and enzymes
small intestine: the organ where digestion of protein, fats, and carbohydrates is completed
stomach: a saclike organ containing acidic digestive juices
vitamin: an organic substance necessary in small amounts to sustain life

11.3 Circulatory and Respiratory Systems

Learning Objectives

By the end of this section, you will be able to:

- Describe the passage of air from the outside environment to the lungs
- Explain how the lungs are protected from particulate matter
- Describe the function of the circulatory system
- Describe the cardiac cycle
- Explain how blood flows through the body

Animals are complex multicellular organisms that require a mechanism for transporting nutrients throughout their bodies and removing wastes. The human circulatory system has a complex network of blood vessels that reach all parts of the body. This extensive network supplies the cells, tissues, and organs with oxygen and nutrients, and removes carbon dioxide and waste compounds.

The medium for transport of gases and other molecules is the blood, which continually circulates through the system. Pressure differences within the system cause the movement of the blood and are created by the pumping of the heart.

Gas exchange between tissues and the blood is an essential function of the circulatory system. In humans, other mammals, and birds, blood absorbs oxygen and releases carbon dioxide in the lungs. Thus the circulatory and respiratory system, whose function is to obtain oxygen and discharge carbon dioxide, work in tandem.

The Respiratory System (Basic level)

Take a breath in and hold it. Wait several seconds and then let it out. Humans, when they are not exerting themselves, breathe approximately 15 times per minute on average. This equates to about 900 breaths an hour or 21,600 breaths per day. With every inhalation, air fills the lungs, and with every exhalation, it rushes back out. That air is doing more than just inflating and deflating the lungs in the chest cavity. The air contains oxygen that crosses the lung tissue, enters the bloodstream, and travels to organs and tissues. There, oxygen is exchanged for carbon dioxide, which is a cellular waste material. Carbon dioxide exits the cells, enters the bloodstream, travels back to the lungs, and is expired out of the body during exhalation.

Breathing is both a voluntary and an involuntary event. How often a breath is taken and how much air is inhaled or exhaled is regulated by the respiratory center in the brain in response to signals it receives

about the carbon dioxide content of the blood. However, it is possible to override this automatic regulation for activities such as speaking, singing and swimming under water.

During inhalation the diaphragm descends creating a negative pressure around the lungs and they begin to inflate, drawing in air from outside the body. The air enters the body through the nasal cavity located just inside the nose (Figure 11.9). As the air passes through the nasal cavity, the air is warmed to body temperature and humidified by moisture from mucous membranes. 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 by hairs, mucus, and cilia. Air is also chemically sampled by the sense of smell.

From the nasal cavity, air passes through the pharynx (throat) and the larynx (voice box) as it makes its way to the trachea (Figure 11.9). The main function of the trachea is to funnel the inhaled air to the lungs and the exhaled air back out of the body. The human trachea is a cylinder, about 25 to 30 cm (9.8–11.8 in) long, which sits in front of the esophagus and extends from the pharynx into the chest cavity to the lungs. It is made of incomplete rings of cartilage and smooth muscle. The cartilage provides strength and support to the trachea to keep the passage open. The trachea is lined with cells that have cilia and secrete mucus. The mucus catches particles that have been inhaled, and the cilia move the particles toward the pharynx.

The end of the trachea divides into two bronchi that enter the right and left lung. Air enters the lungs through the primary bronchi. The primary bronchus divides, creating smaller and smaller diameter bronchi until the passages are under 1 mm (.03 in) in diameter when they are called bronchioles as they split and spread through the lung. Like the trachea, the bronchus and bronchioles are made of cartilage and smooth muscle. Bronchi are innervated by nerves of both the parasympathetic and sympathetic nervous systems that control muscle contraction (parasympathetic) or relaxation (sympathetic) in the bronchi and bronchioles, depending on the nervous system's cues. The final bronchioles are the respiratory bronchioles. Alveolar ducts are attached to the end of each respiratory bronchiole. At the end of each duct are alveolar sacs, each containing 20 to 30 alveoli. Gas exchange occurs only in the alveoli. The alveoli are thin-walled and look like tiny bubbles within the sacs. The alveoli are in direct contact with capillaries of the circulatory system. Such intimate contact ensures that oxygen will diffuse from the alveoli into the blood. In addition, carbon dioxide will diffuse from the blood into the alveoli to be exhaled. The anatomical arrangement of capillaries and alveoli emphasizes the structural and functional relationship of the respiratory and circulatory systems. Estimates for the surface area of alveoli in the lungs vary around 100 m^2. This large area is about the area of half a tennis court. This large surface area, combined with the thin-walled nature of the alveolar cells, allows gases to easily diffuse across the cells.

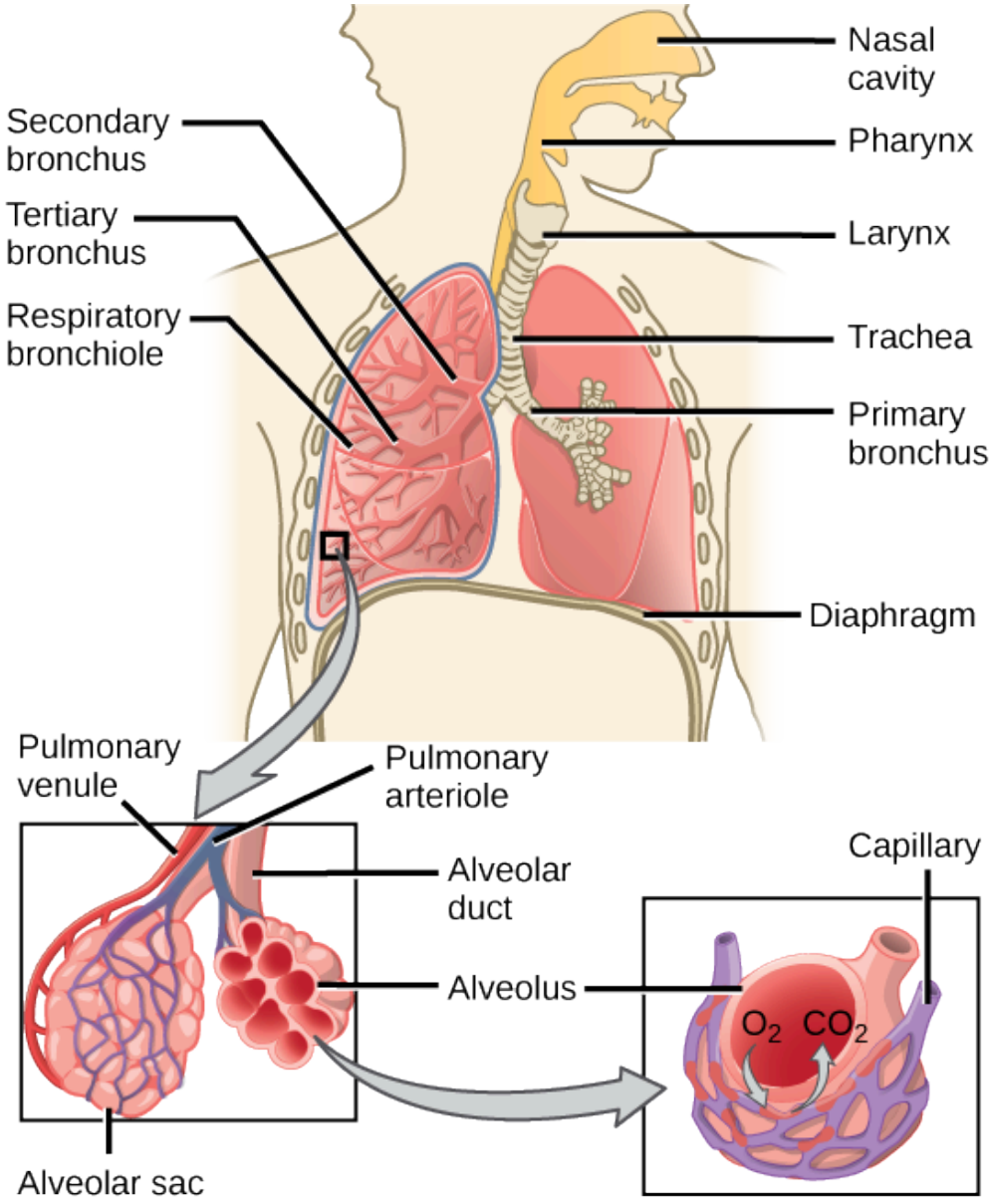


Figure 11.9 Air enters the respiratory system through the nasal cavity, and then passes through the pharynx and the trachea into the lungs. (credit: modification of work by NCI)

Systems of Gas Exchange

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

All aerobic organisms require oxygen to carry out their metabolic functions. Along the evolutionary tree, different organisms have devised different means of obtaining oxygen from the surrounding atmosphere. The environment in which the animal lives greatly determines how an animal respires. The complexity of the respiratory system is correlated with the size of the organism. As animal size

increases, diffusion distances increase and the ratio of surface area to volume drops. In unicellular organisms, diffusion across the cell membrane is sufficient for supplying oxygen to the cell (Figure 11.10). Diffusion is a slow, passive transport process. In order for diffusion to be a feasible means of providing oxygen to the cell, the rate of oxygen uptake must match the rate of diffusion across the membrane. In other words, if the cell were very large or thick, diffusion would not be able to provide oxygen quickly enough to the inside of the cell. Therefore, dependence on diffusion as a means of obtaining oxygen and removing carbon dioxide remains feasible only for small organisms or those with highly-flattened bodies, such as many flatworms (Platyhelminthes). Larger organisms had to evolve specialized respiratory tissues, such as gills, lungs, and respiratory passages accompanied by a complex circulatory systems, to transport oxygen throughout their entire body.

Figure 11.10 The cell of the unicellular algae Ventricaria ventricosa is one of the largest known, reaching one to five centimeters in diameter. Like all single-celled organisms, V. ventricosa exchanges gases across the cell membrane.

Direct Diffusion

For small multicellular organisms, diffusion across the outer membrane is sufficient to meet their oxygen needs. Gas exchange by direct diffusion across surface membranes is efficient for organisms less than 1 mm in diameter. In simple organisms, such as cnidarians and flatworms, every cell in the body is close to the external environment. Their cells are kept moist and gases diffuse quickly via direct diffusion. Flatworms are small, literally flat worms, which 'breathe' through diffusion across the outer membrane (Figure 11.11). The flat shape of these organisms increases the surface area for diffusion, ensuring that each cell within the body is close to the outer membrane surface and has access to oxygen. If the flatworm had a cylindrical body, then the cells in the center would not be able to get oxygen.

Figure 11.11. 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 cell membranes.

Organisms that live in water need to obtain oxygen from the water. Oxygen dissolves in water but at a lower concentration than in the atmosphere. The atmosphere has roughly 21 percent oxygen. In water, the oxygen concentration is much smaller than that. Fish and many other aquatic organisms have evolved gills to take up the dissolved oxygen from water (Figure 11.12). 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 11.12.
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 11.13. Similarly, carbon dioxide molecules in the blood diffuse from the blood (high concentration) to water (low concentration).

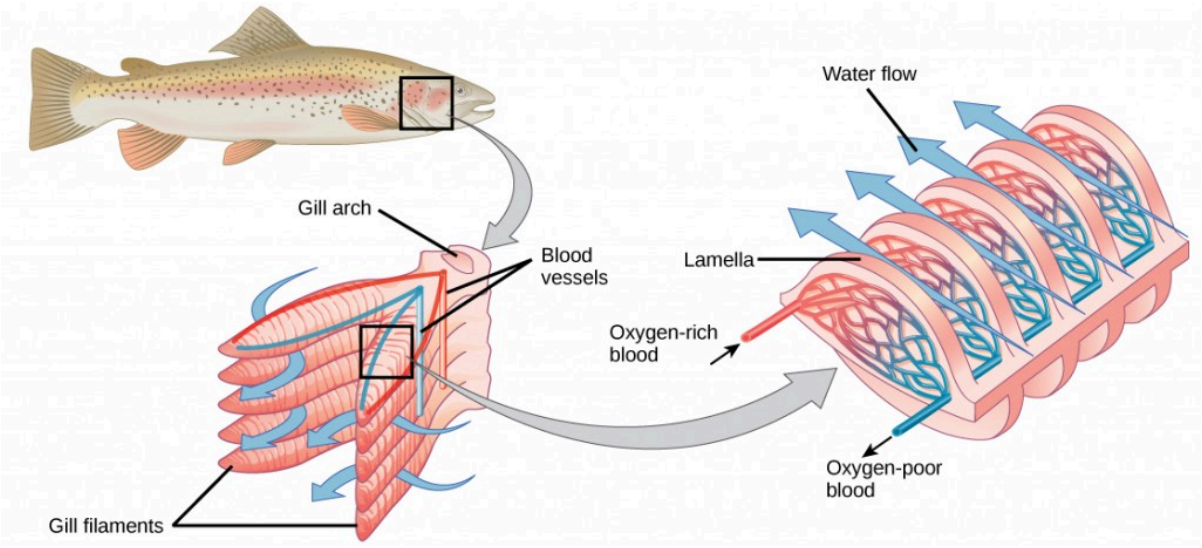


Figure 11.13. As water flows over the gills, oxygen is transferred to blood via the veins. (credit "fish": modification of work by Duane Raver, NOAA)

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. The tubes in the tracheal system are made of a polymeric material called chitin.

Insect bodies have openings, called spiracles, along the thorax and abdomen. These openings connect to the tubular network, allowing oxygen to pass into the body (Figure 11.14) 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.

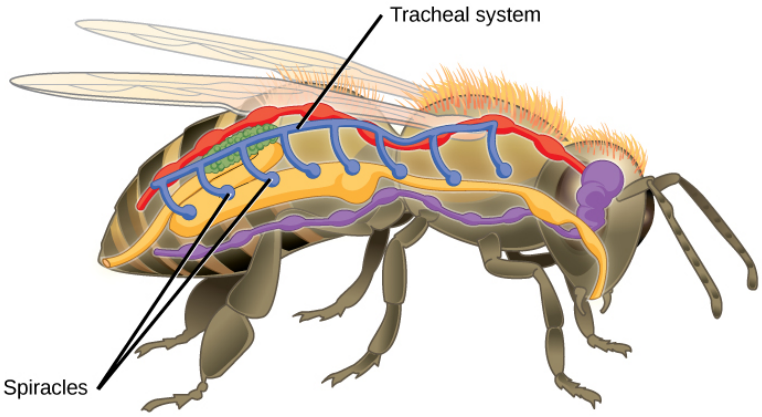


Figure 11.14. Insects perform respiration via a tracheal system.

Mammalian Systems

In mammals, pulmonary ventilation occurs via inhalation (breathing). During inhalation, air enters the body through the**nasal cavity** located just inside the nose (Figure 11.15). 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.

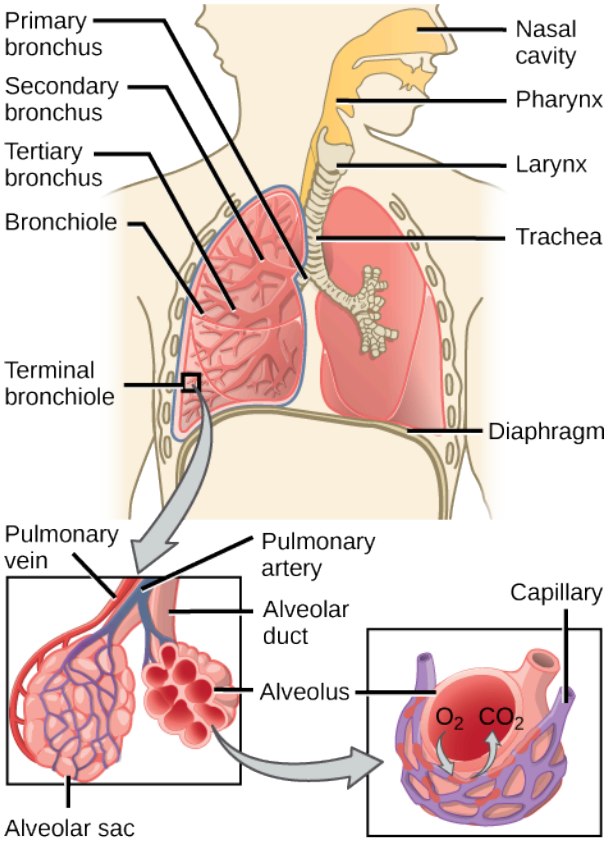


Figure 11.15. Air enters the respiratory system through the nasal cavity and pharynx, and then passes through the trachea and into the bronchi, which bring air into the lungs. (credit: modification of work by NCI)

Which of the following statements about the mammalian respiratory system is false?

1. When we breathe in, air travels from the pharynx to the trachea.
2. The bronchioles branch into bronchi.
3. Alveolar ducts connect to alveolar sacs.
4. Gas exchange between the lung and blood takes place in the alveolus.

From the nasal cavity, air passes through the **pharynx** (throat) and the **larynx** (voice box), as it makes its way to the **trachea** (Figure 11.16). 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. It is made of incomplete rings of hyaline cartilage and smooth muscle (Figure 11.17). 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.

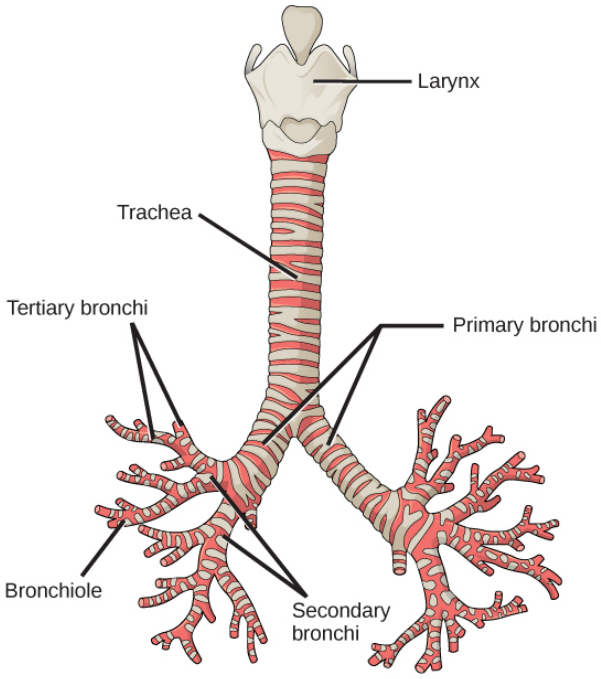


Figure 11.16.
The trachea and bronchi are made of incomplete rings of cartilage. (credit: modification of work by Gray's Anatomy)

Lungs: Bronchi and Alveoli

The end of the trachea bifurcates (divides) to the right and left lungs. The lungs are not identical. The right lung is larger and contains three lobes, whereas the smaller left lung contains two lobes (Figure 11.17). The muscular **diaphragm,** which facilitates breathing, is inferior (below) to the lungs and marks the end of the thoracic cavity.

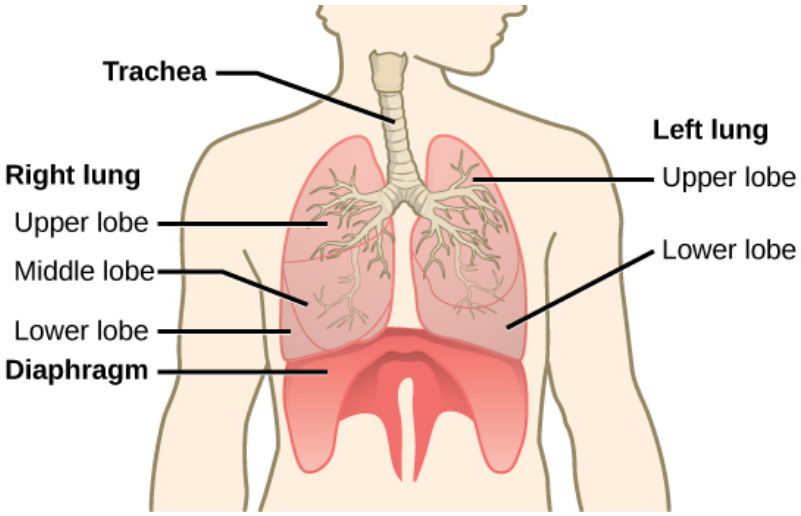


Figure 11.17. The trachea bifurcates into the right and left bronchi in the lungs. The right lung is made of three lobes and is larger. To accommodate the heart, the left lung is smaller and has only two lobes.

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

The **terminal bronchioles** subdivide into microscopic branches called respiratory bronchioles. The respiratory bronchioles subdivide into several alveolar ducts. Numerous alveoli and alveolar sacs surround the alveolar ducts. The alveolar sacs resemble bunches of grapes tethered to the end of the bronchioles (Figure 11.18). In the acinar region, the**alveolar ducts** are attached to the end of each bronchiole. At the end of each duct are approximately 100 **alveolar sacs,** each containing 20 to 30 **alveoli** that are 200 to 300 microns in diameter. Gas exchange occurs only in alveoli. Alveoli are made of thin-walled parenchymal cells, typically one-cell thick, that look like tiny bubbles within the sacs. Alveoli are in direct contact with capillaries (one-cell thick) of the circulatory system. Such intimate contact ensures that oxygen will diffuse from alveoli into the blood and be distributed to the cells of the body. In addition, the carbon dioxide that was produced by cells as a waste product will diffuse from the blood into alveoli to be exhaled. The anatomical arrangement of capillaries and alveoli emphasizes the structural and functional relationship of the respiratory and circulatory systems. Because there are so many alveoli (~300 million per lung) within each alveolar sac and so many sacs at the end of each alveolar duct, the lungs have a sponge-like consistency. This organization produces a very large surface area that is available for gas exchange. The surface area of alveoli in the lungs is approximately 75 m^2. This large surface area, combined with the thin-walled nature of the alveolar parenchymal cells, allows gases to easily diffuse across the cells.

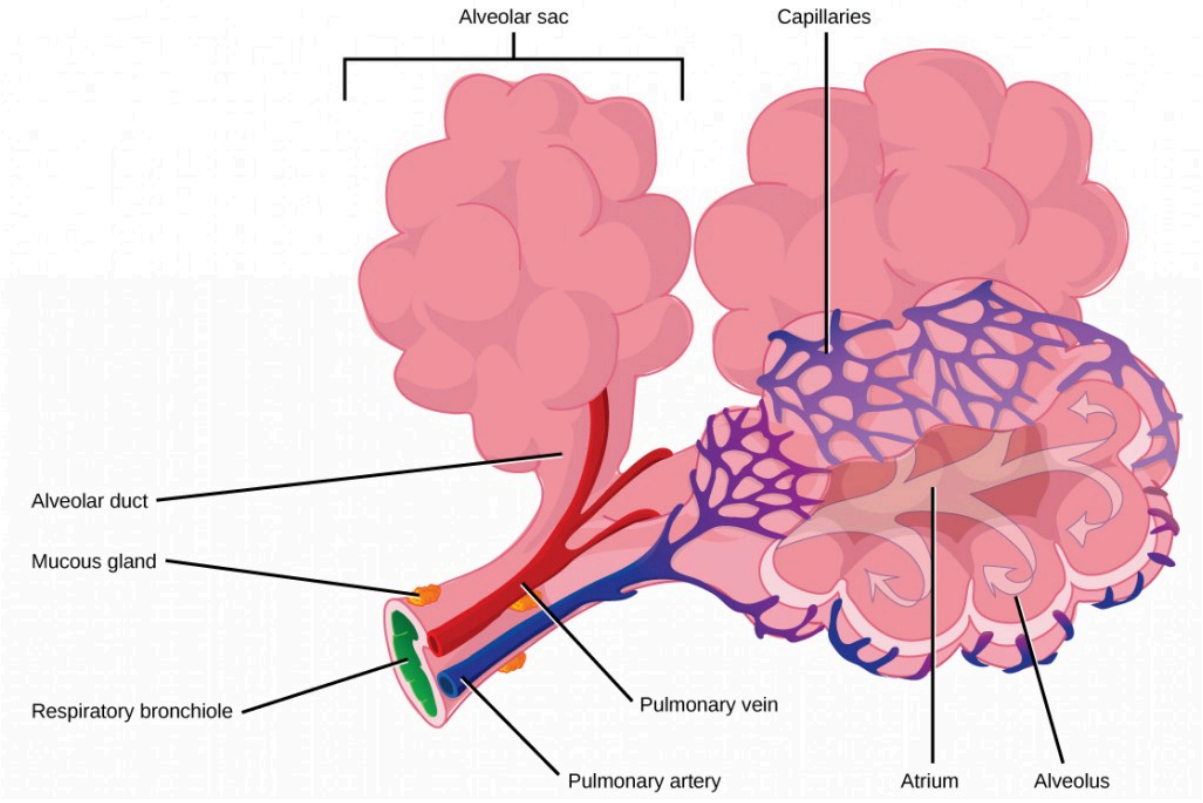


Figure 11.18.
Terminal bronchioles are connected by respiratory bronchioles to alveolar ducts and alveolar sacs. Each alveolar sac contains 20 to 30 spherical alveoli and has the appearance of a bunch of grapes. Air flows into the atrium of the alveolar sac, then circulates into alveoli where gas exchange occurs with the capillaries. Mucous glands secrete mucous into the airways, keeping them moist and flexible. (credit: modification of work by Mariana Ruiz Villareal)

Concept in Action

Watch the following video to review the respiratory system.

Protective Mechanisms

The air that organisms breathe contains **particulate matter** such as dust, dirt, viral particles, and bacteria that can damage the lungs or trigger allergic immune responses. The respiratory system contains several protective mechanisms to avoid problems or tissue damage. In the nasal cavity, hairs and mucus trap small particles, viruses, bacteria, dust, and dirt to prevent their entry.

If particulates do make it beyond the nose, or enter through the mouth, the bronchi and bronchioles of the lungs also contain several protective devices. The lungs produce **mucus**—a sticky substance made of **mucin**, a complex glycoprotein, as well as salts and water—that traps particulates. The bronchi and bronchioles contain cilia, small hair-like projections that line the walls of the bronchi and bronchioles (Figure 11.19). These cilia beat in unison and move mucus and particles out of the bronchi and bronchioles back up to the throat where it is swallowed and eliminated via the esophagus.

In humans, for example, tar and other substances in cigarette smoke destroy or paralyze the cilia, making the removal of particles more difficult. In addition, smoking causes the lungs to produce more mucus, which the damaged cilia are not able to move. This causes a persistent cough, as the lungs try to rid themselves of particulate matter, and makes smokers more susceptible to respiratory ailments.

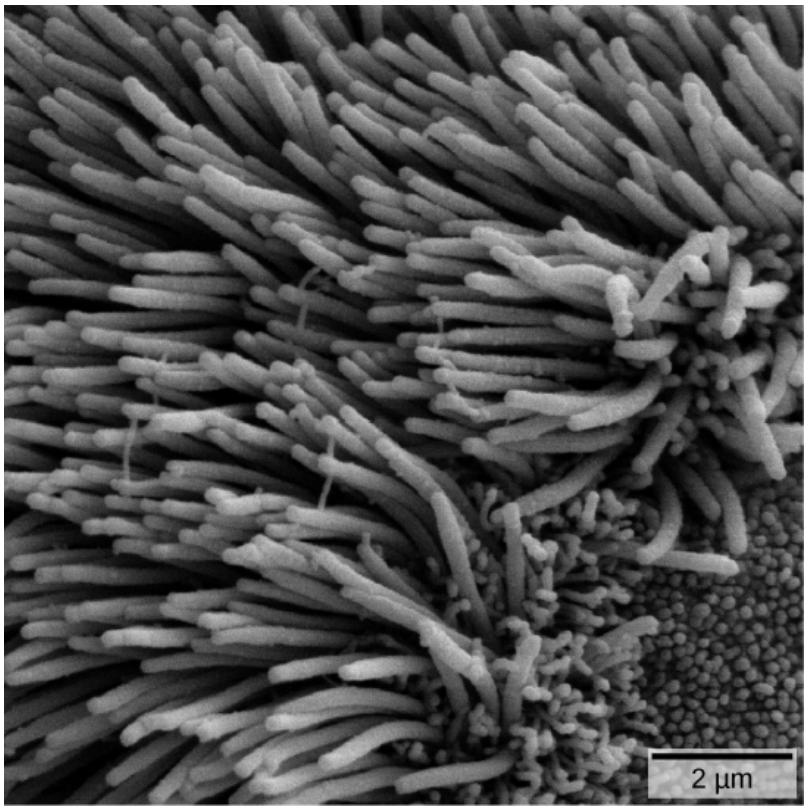


Figure 11.19.
The bronchi and bronchioles contain cilia that help move mucus and other particles out of the lungs. (credit: Louisa Howard, modification of work by Dartmouth Electron Microscope Facility)

Summary

Animal respiratory systems are designed to facilitate gas exchange. In mammals, air is warmed and humidified in the nasal cavity. Air then travels down the pharynx, through the trachea, and into the lungs. In the lungs, air passes through the branching bronchi, reaching the respiratory bronchioles, which house the first site of gas exchange. The respiratory bronchioles open into the alveolar ducts, alveolar sacs, and alveoli. Because there are so many alveoli and alveolar sacs in the lung, the surface area for gas exchange is very large. Several protective mechanisms are in place to prevent damage or infection. These include the hair and mucus in the nasal cavity that trap dust, dirt, and other particulate

matter before they can enter the system. In the lungs, particles are trapped in a mucus layer and transported via cilia up to the esophageal opening at the top of the trachea to be swallowed.

The Circulatory System

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

The Heart

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

The heart is asymmetrical, with the left side being larger than the right side, correlating with the different sizes of the pulmonary and systemic circuits (Figure 11.10). In humans, the heart is about the size of a clenched fist; it is divided into four chambers: two atria and two ventricles. There is one atrium and one ventricle on the right side and one atrium and one ventricle on the left side. The right atrium receives deoxygenated blood from the systemic circulation through the major veins: the superior vena cava, which drains blood from the head and from the veins that come from the arms, as well as the inferior vena cava, which drains blood from the veins that come from the lower organs and the legs. This deoxygenated blood then passes to the right ventricle through the tricuspid valve, which prevents the backflow of blood. After it is filled, the right ventricle contracts, pumping the blood to the lungs for reoxygenation. The left atrium receives the oxygen-rich blood from the lungs. This blood passes through the bicuspid valve to the left ventricle where the blood is pumped into the aorta. The aorta is the major artery of the body, taking oxygenated blood to the organs and muscles of the body. This pattern of pumping is referred to as double circulation and is found in all mammals. (Figure 11.20).

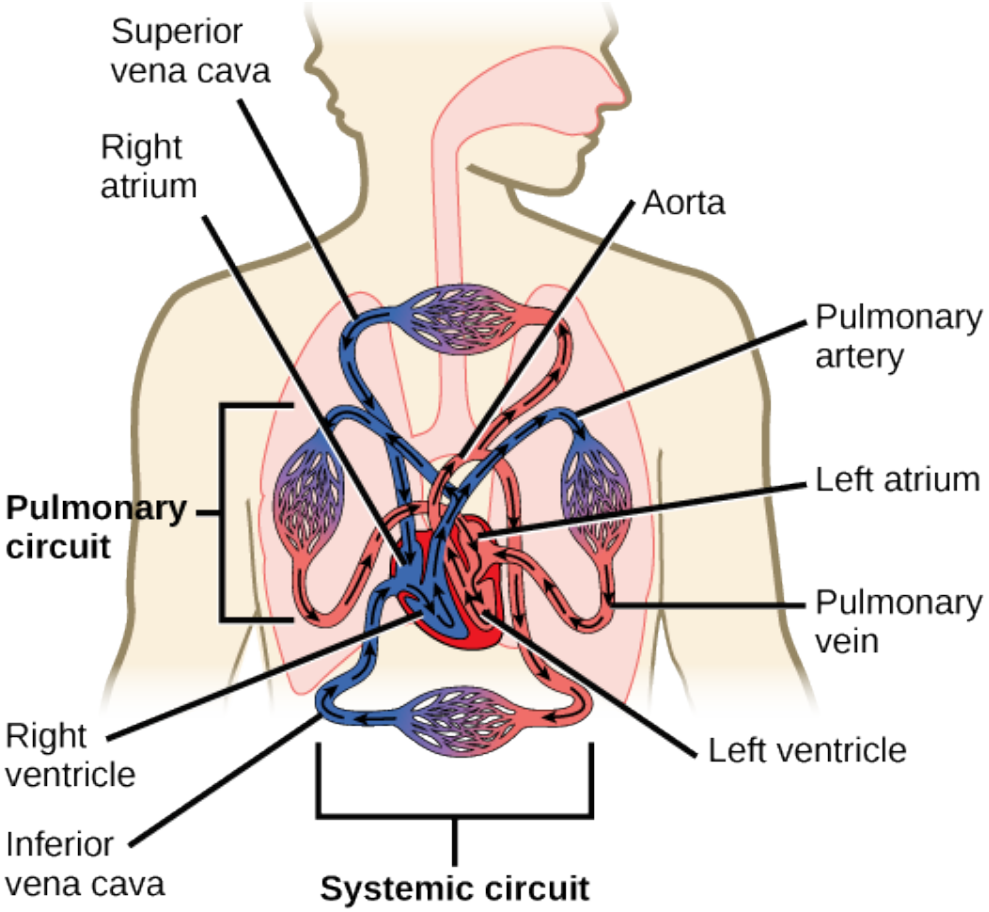


Figure 11.20 The heart is divided into four chambers, two atria, and two ventricles. Each chamber is separated by one-way valves. The right side of the heart receives deoxygenated blood from the body and pumps it to the lungs. The left side of the heart pumps blood to the rest of the body.

The Cardiac Cycle

The main purpose of the heart is to pump blood through the body; it does so in a repeating sequence called the cardiac cycle. The cardiac cycle is the flow of blood through the heart coordinated by electrochemical signals that cause the heart muscle to contract and relax. In each cardiac cycle, a sequence of contractions pushes out the blood, pumping it through the body; this is followed by a relaxation phase, where the heart fills with blood. These two phases are called the systole (contraction) and diastole (relaxation), respectively (Figure 11.21). The signal for contraction begins at a location on the outside of the right atrium. The electrochemical signal moves from there across the atria causing them to contract. The contraction of the atria forces blood through the valves into the ventricles. Closing of these valves caused by the contraction of the ventricles produces a "lub" sound. The signal has, by this time, passed down the walls of the heart, through a point between the right atrium and right ventricle. The signal then causes the ventricles to contract. The ventricles contract together forcing blood into the aorta and the pulmonary arteries. Closing of the valves to these arteries caused by blood being drawn back toward the heart during ventricular relaxation produces a monosyllabic "dub" sound.

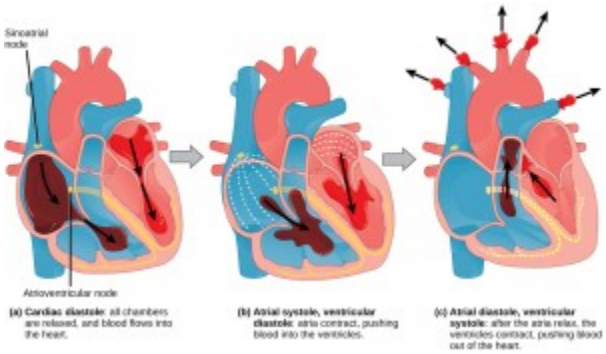


Figure 11.21 In each cardiac cycle, a series of contractions (systoles) and relaxations (diastoles) pumps blood through the heart and through the body. (a) During cardiac diastole, blood flows into the heart while all chambers are relaxed. (b) Then the ventricles remain relaxed while atrial systole pushes blood into the ventricles. (c) Once the atria relax again, ventricle systole pushes blood out of the heart.

The pumping of the heart is a function of the cardiac muscle cells, or cardiomyocytes, that make up the heart muscle. Cardiomyocytes are distinctive muscle cells that are striated like skeletal muscle but pump rhythmically and involuntarily like smooth muscle; adjacent cells are connected by intercalated disks found only in cardiac muscle. These connections allow the electrical signal to travel directly to neighboring muscle cells.

The electrical impulses in the heart produce electrical currents that flow through the body and can be measured on the skin using electrodes. This information can be observed as an electrocardiogram (ECG) a recording of the electrical impulses of the cardiac muscle.

Concept in Action

Visit the following website to see the heart's pacemaker, or electrocardiogram system, in action.

Blood Vessels

The blood from the heart is carried through the body by a complex network of blood vessels (Figure 11.22). Arteries take blood away from the heart. The main artery of the systemic circulation is the aorta; it branches into major arteries that take blood to different limbs and organs. The aorta and

arteries near the heart have heavy but elastic walls that respond to and smooth out the pressure differences caused by the beating heart. Arteries farther away from the heart have more muscle tissue in their walls that can constrict to affect flow rates of blood. The major arteries diverge into minor arteries, and then smaller vessels called arterioles, to reach more deeply into the muscles and organs of the body.

Arterioles diverge into capillary beds. Capillary beds contain a large number, 10's to 100's of capillaries that branch among the cells of the body. Capillaries are narrow-diameter tubes that can fit single red blood cells and are the sites for the exchange of nutrients, waste, and oxygen with tissues at the cellular level. Fluid also leaks from the blood into the interstitial space from the capillaries. The capillaries converge again into venules that connect to minor veins that finally connect to major veins. Veins are blood vessels that bring blood high in carbon dioxide back to the heart. Veins are not as thick-walled as arteries, since pressure is lower, and they have valves along their length that prevent backflow of blood away from the heart. The major veins drain blood from the same organs and limbs that the major arteries supply.

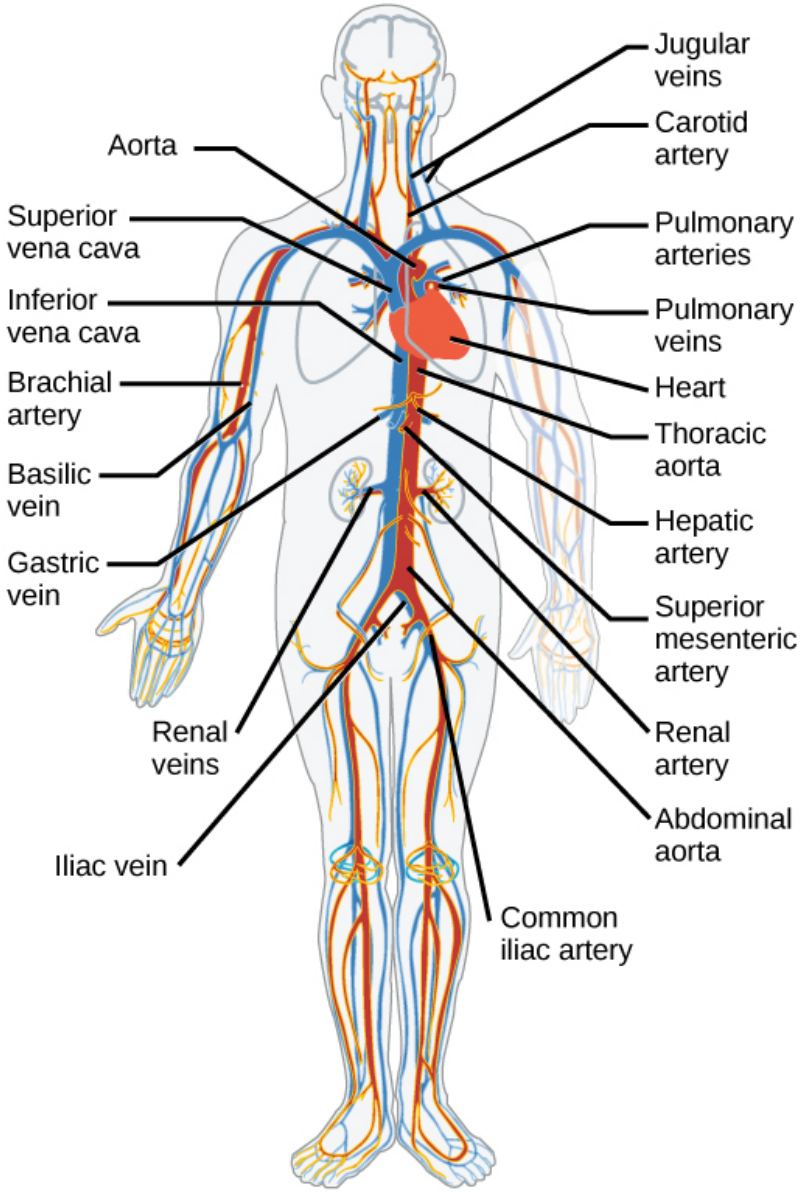


Figure 11.22 The arteries of the body, indicated in red, start at the aortic arch and branch to supply the organs and muscles of the body with oxygenated blood. The veins of the body, indicated in blue, return blood to the heart. The pulmonary arteries are blue to reflect the fact that they are deoxygenated, and the pulmonary veins are red to reflect that they are oxygenated. (credit: modification of work by Mariana Ruiz Villareal)

Section Summary

Animal respiratory systems are designed to facilitate gas exchange. In mammals, air is warmed and humidified in the nasal cavity. Air then travels down the pharynx and larynx, through the trachea, and into the lungs. In the lungs, air passes through the branching bronchi, reaching the respiratory bronchioles. The respiratory bronchioles open up into the alveolar ducts, alveolar sacs, and alveoli.

Because there are so many alveoli and alveolar sacs in the lung, the surface area for gas exchange is very large.

The mammalian circulatory system is a closed system with double circulation passing through the lungs and the body. It consists of a network of vessels containing blood that circulates because of pressure differences generated by the heart.

The heart contains two pumps that move blood through the pulmonary and systemic circulations. There is one atrium and one ventricle on the right side and one atrium and one ventricle on the left side. The pumping of the heart is a function of cardiomyocytes, distinctive muscle cells that are striated like skeletal muscle but pump rhythmically and involuntarily like smooth muscle. The signal for contraction begins in the wall of the right atrium. The electrochemical signal causes the two atria to contract in unison; then the signal causes the ventricles to contract. The blood from the heart is carried through the body by a complex network of blood vessels; arteries take blood away from the heart, and veins bring blood back to the heart.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here: https://opentextbc.ca/biology/?p=4667#h5p-72

Glossary

alveolus: (plural: alveoli) (also, air sacs) the terminal structure of the lung passage where gas exchange occurs
aorta: the major artery that takes blood away from the heart to the systemic circulatory system
artery: a blood vessel that takes blood away from the heart
atrium: (plural: atria) a chamber of the heart that receives blood from the veins
bicuspid valve: a one-way opening between the atrium and the ventricle in the left side of the heart
bronchi: (singular: bronchus) smaller branches of cartilaginous tissue that stem off of the trachea; air is funneled through the bronchi to the region where gas exchange occurs in the alveoli
bronchiole: an airway that extends from the main bronchus to the alveolar sac
capillary: the smallest blood vessel that allows the passage of individual blood cells and the site of diffusion of oxygen and nutrient exchange
cardiac cycle: the filling and emptying the heart of blood caused by electrical signals that cause the heart muscles to contract and relax
closed circulatory system: a system that has the blood separated from the bodily interstitial fluid and contained in blood vessels
diaphragm: a skeletal muscle located under lungs that encloses the lungs in the thorax

diastole: the relaxation phase of the cardiac cycle when the heart is relaxed and the ventricles are filling with blood
electrocardiogram (ECG): a recording of the electrical impulses of the cardiac muscle
inferior vena cava: the major vein of the body returning blood from the lower parts of the body to the right atrium
larynx: the voice box, located within the throat
nasal cavity: an opening of the respiratory system to the outside environment
open circulatory system: a circulatory system that has the blood mixed with interstitial fluid in the body cavity and directly bathes the organs
pharynx: the throat
primary bronchus: (also, main bronchus) a region of the airway within the lung that attaches to the trachea and bifurcates to form the bronchioles
pulmonary circulation: the flow of blood away from the heart through the lungs where oxygenation occurs and then back to the heart
superior vena cava: the major vein of the body returning blood from the upper part of the body to the right atrium
systemic circulation: the flow of blood away from the heart to the brain, liver, kidneys, stomach, and other organs, the limbs, and the muscles of the body, and then back to the heart
systole: the contraction phase of cardiac cycle when the ventricles are pumping blood into the arteries
trachea: the cartilaginous tube that transports air from the throat to the lungs
tricuspid valve: a one-way opening between the atrium and the ventricle in the right side of the heart
vein: a blood vessel that brings blood back to the heart
ventricle: (of the heart) a large chamber of the heart that pumps blood into arteries

11.4 Endocrine System

Learning Objectives

By the end of this section, you will be able to:

- List the different types of hormones and explain their roles in maintaining homeostasis
- Explain how hormones work
- Explain how hormone production is regulated
- Describe the role of different glands in the endocrine system
- Explain how the different glands work together to maintain homeostasis

The endocrine system produces hormones that function to control and regulate many different body processes. The endocrine system coordinates with the nervous system to control the functions of the other organ systems. Cells of the endocrine system produce molecular signals called hormones. These cells may compose endocrine glands, may be tissues or may be located in organs or tissues that have functions in addition to hormone production. Hormones circulate throughout the body and stimulate a response in cells that have receptors able to bind with them. The changes brought about in the receiving cells affect the functioning of the organ system to which they belong. Many of the hormones are secreted in response to signals from the nervous system, thus the two systems act in concert to effect changes in the body.

Hormones

Maintaining homeostasis within the body requires the coordination of many different systems and organs. One mechanism of communication between neighboring cells, and between cells and tissues in distant parts of the body, occurs through the release of chemicals called hormones. Hormones are released into body fluids, usually blood, which carries them to their target cells where they elicit a response. The cells that secrete hormones are often located in specific organs, called endocrine glands, and the cells, tissues, and organs that secrete hormones make up the endocrine system. Examples of endocrine organs include the pancreas, which produces the hormones insulin and glucagon to regulate blood-glucose levels, 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.

The endocrine glands differ from the exocrine glands. Exocrine glands secrete chemicals through ducts that lead outside the gland (not to the blood). For example, sweat produced by sweat glands is released into ducts that carry sweat to the surface of the skin. The pancreas has both endocrine and exocrine

functions because besides releasing hormones into the blood. It also produces digestive juices, which are carried by ducts into the small intestine.

Endocrinologist

An endocrinologist is a medical doctor who specializes in treating endocrine disorders. An endocrine surgeon specializes in the surgical treatment of endocrine diseases and glands. Some of the diseases that are managed by endocrinologists include disorders of the pancreas (diabetes mellitus), disorders of the pituitary (gigantism, acromegaly, and pituitary dwarfism), disorders of the thyroid gland (goiter and Graves' disease), and disorders of the adrenal glands (Cushing's disease and Addison's disease).

Endocrinologists are required to assess patients and diagnose endocrine disorders through extensive use of laboratory tests. Many endocrine diseases are diagnosed using tests that stimulate or suppress endocrine organ functioning. Blood samples are then drawn to determine the effect of stimulating or suppressing an endocrine organ on the production of hormones. For example, to diagnose diabetes mellitus, patients are required to fast for 12 to 24 hours. They are then given a sugary drink, which stimulates the pancreas to produce insulin to decrease blood-glucose levels. A blood sample is taken one to two hours after the sugar drink is consumed. If the pancreas is functioning properly, the blood-glucose level will be within a normal range. Another example is the A1C test, which can be performed during blood screening. The A1C test measures average blood-glucose levels over the past two to three months. The A1C test is an indicator of how well blood glucose is being managed over a long time.

Once a disease such as diabetes has been diagnosed, endocrinologists can prescribe lifestyle changes and medications to treat the disease. Some cases of diabetes mellitus can be managed by exercise, weight loss, and a healthy diet; in other cases, medications may be required to enhance insulin's production or effect. If the disease cannot be controlled by these means, the endocrinologist may prescribe insulin injections.

In addition to clinical practice, endocrinologists may also be involved in primary research and development activities. For example, ongoing islet transplant research is investigating how healthy pancreas islet cells may be transplanted into diabetic patients. Successful islet transplants may allow patients to stop taking insulin injections.

How Hormones Work

Hormones cause changes in target cells by binding to specific cell-surface or intracellular hormone receptors, molecules embedded in the cell membrane or floating in the cytoplasm with a binding site that matches a binding site on the hormone molecule. 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 or in many different cells or may be limited to a small number of specialized cells. For example, thyroid hormones act on many different tissue types, stimulating metabolic activity throughout the body. Cells can have many receptors for the same hormone but often also possess receptors for different types of hormones. The number of receptors that respond to a hormone determines the cell's sensitivity to that hormone, and the resulting cellular response. Additionally, the number of receptors available to respond to a hormone can change over time, resulting in increased or decreased cell sensitivity. In up-regulation, the

number of receptors increases in response to rising hormone levels, making the cell more sensitive to the hormone and allowing for more cellular activity. When the number of receptors decreases in response to rising hormone levels, called down-regulation, cellular activity is reduced.

Endocrine Glands

The endocrine glands secrete hormones into the surrounding interstitial fluid; those hormones then diffuse into blood and are carried to various organs and tissues within the body. The endocrine glands include the pituitary, thyroid, parathyroid, adrenal glands, gonads, pineal, and pancreas.

The pituitary gland, sometimes called the hypophysis, is located at the base of the brain (Figure 11.23 **a**). It is attached to the hypothalamus. The posterior lobe stores and releases oxytocin and antidiuretic hormone produced by the hypothalamus. The anterior lobe responds to hormones produced by the hypothalamus by producing its own hormones, most of which regulate other hormone-producing glands.

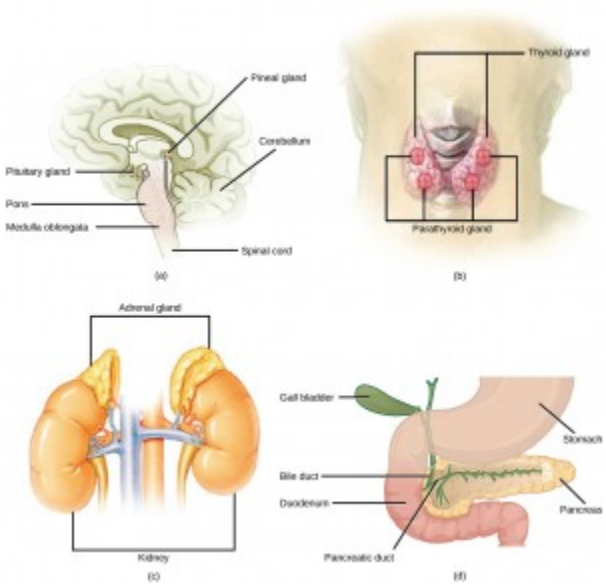


Figure 11.23 (a) The pituitary gland sits at the base of the brain, just above the brain stem. (b) The parathyroid glands are located on the posterior of the thyroid gland. (c) The adrenal glands are on top of the kidneys. d) The pancreas is found between the stomach and the small intestine. (credit: modification of work by NCI, NIH)

The anterior pituitary produces six hormones: growth hormone, prolactin, thyroid-stimulating hormone, adrenocorticotropic hormone, follicle-stimulating hormone, and luteinizing hormone. Growth hormone stimulates cellular activities like protein synthesis that promote growth. Prolactin stimulates the production of milk by the mammary glands. The other hormones produced by the anterior pituitary regulate the production of hormones by other endocrine tissues (Table 11.1). The posterior pituitary is significantly different in structure from the anterior pituitary. It is a part of the brain, extending down

from the hypothalamus, and contains mostly nerve fibers that extend from the hypothalamus to the posterior pituitary.

The thyroid gland is located in the neck, just below the larynx and in front of the trachea (Figure 11.23 **b**). It is a butterfly-shaped gland with two lobes that are connected. The 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. T_3 and T_4 are released by the thyroid in response to thyroid-stimulating hormone produced by the anterior pituitary, and both T3 and T4 have the effect of stimulating metabolic activity in the body and increasing energy use. A third hormone, calcitonin, is also produced by the thyroid. Calcitonin is released in response to rising calcium ion concentrations in the blood and has the effect of reducing those levels.

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 (Figure 11.23 **b**).

The parathyroid glands produce parathyroid hormone. Parathyroid hormone increases blood calcium concentrations when calcium ion levels fall below normal.

The adrenal glands are located on top of each kidney (Figure 11.23 **c**). The adrenal glands consist of an outer adrenal cortex and an inner adrenal medulla. These regions secrete different hormones.

The adrenal cortex produces mineralocorticoids, glucocorticoids, and androgens. The main mineralocorticoid is aldosterone, which regulates the concentration of ions in urine, sweat, 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 glucocorticoids maintain proper blood-glucose levels between meals. They also control a response to stress by increasing glucose synthesis from fats and proteins and interact with epinephrine to cause vasoconstriction. Androgens are sex hormones that are produced in small amounts by the adrenal cortex. They do not normally affect sexual characteristics and may supplement sex hormones released from the gonads. The adrenal medulla contains two types of secretory cells: one that produces epinephrine (adrenaline) and another that produces norepinephrine (noradrenaline). Epinephrine and norepinephrine cause immediate, short-term changes in response to stressors, inducing the so-called fight-or-flight response. The responses include increased heart rate, breathing rate, cardiac muscle contractions, and blood-glucose levels. They also accelerate the breakdown of glucose in skeletal muscles and stored fats in adipose tissue, and redirect blood flow toward skeletal muscles and away from skin and viscera. The release of epinephrine and norepinephrine is stimulated by neural impulses from the sympathetic nervous system that originate from the hypothalamus.

The pancreas is an elongate organ located between the stomach and the proximal portion of the small intestine (Figure 11.23 **d**). It contains both exocrine cells that excrete digestive enzymes and endocrine cells that release hormones.

The endocrine cells of the pancreas form clusters called pancreatic islets or the islets of Langerhans. Among the cell types in each pancreatic islet are the alpha cells, which produce the hormone glucagon, and the beta cells, which produce the hormone insulin. These hormones regulate blood-glucose levels. Alpha cells release glucagon as blood-glucose levels decline. When blood-glucose levels rise, beta cells release insulin. Glucagon causes the release of glucose to the blood from the liver, and insulin facilitates the uptake of glucose by the body's cells.

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 estrogen and progesterone, which cause secondary sex characteristics, regulate production of eggs, control pregnancy, and prepare the body for childbirth.

There are several organs whose primary functions are non-endocrine but that also possess endocrine functions. These include the heart, kidneys, intestines, thymus, and adipose tissue. The heart has endocrine cells in the walls of the atria that release a hormone in response to increased blood volume. It causes a reduction in blood volume and blood pressure, and reduces the concentration of Na^+ in the blood.

The gastrointestinal tract produces several hormones that aid in digestion. The endocrine cells are located in the mucosa of the GI tract throughout the stomach and small intestine. They trigger the release of gastric juices, which help to break down and digest food in the GI tract.

The kidneys also possess endocrine function. Two of these hormones regulate ion concentrations and blood volume or pressure. Erythropoietin (EPO) is released by kidneys in response to low oxygen levels. EPO triggers the formation of red blood cells in the bone marrow. EPO has been used by athletes to improve performance. But EPO doping has its risks, since it thickens the blood and increases strain on the heart; it also increases the risk of blood clots and therefore heart attacks and stroke.

The thymus is found behind the sternum. The thymus produces hormones referred to as thymosins, which contribute to the development of the immune response in infants. Adipose tissue, or fat tissue, produces the hormone leptin in response to food intake. Leptin produces a feeling of satiety after eating, reducing the urge for further eating.

Table 11.1 Endocrine Glands and Their Associated Hormones

Endocrine Gland	Associated Hormones	Effect
Pituitary (anterior)	growth hormone	promotes growth of body tissues
	prolactin	promotes milk production
	thyroid-stimulating hormone	stimulates thyroid hormone release
	adrenocorticotropic hormone	stimulates hormone release by adrenal cortex
	follicle-stimulating hormone	stimulates gamete production
	luteinizing hormone	stimulates androgen production by gonads in males; stimulates ovulation and production of estrogen and progesterone in females
Pituitary (posterior)	antidiuretic hormone	stimulates water reabsorption by kidneys
	oxytocin	stimulates uterine contractions during childbirth
Thyroid	thyroxine, triiodothyronine	stimulate metabolism
	calcitonin	reduces blood Ca^{2+} levels
Parathyroid	parathyroid hormone	increases blood Ca^{2+} levels
Adrenal (cortex)	aldosterone	increases blood Na^{+} levels
	cortisol, corticosterone, cortisone	increase blood-glucose levels
Adrenal (medulla)	epinephrine, norepinephrine	stimulate fight-or-flight response
Pancreas	insulin	reduces blood-glucose levels
	glucagon	increases blood-glucose levels

Regulation of Hormone Production

Hormone production and release are primarily controlled by negative feedback, as described in the discussion on homeostasis. In this way, the concentration of hormones in blood is maintained within a narrow range. For example, the anterior pituitary signals the thyroid to release thyroid hormones. Increasing levels of these hormones in the blood then give feedback to the hypothalamus and anterior pituitary to inhibit further signaling to the thyroid gland (Figure 11.24).

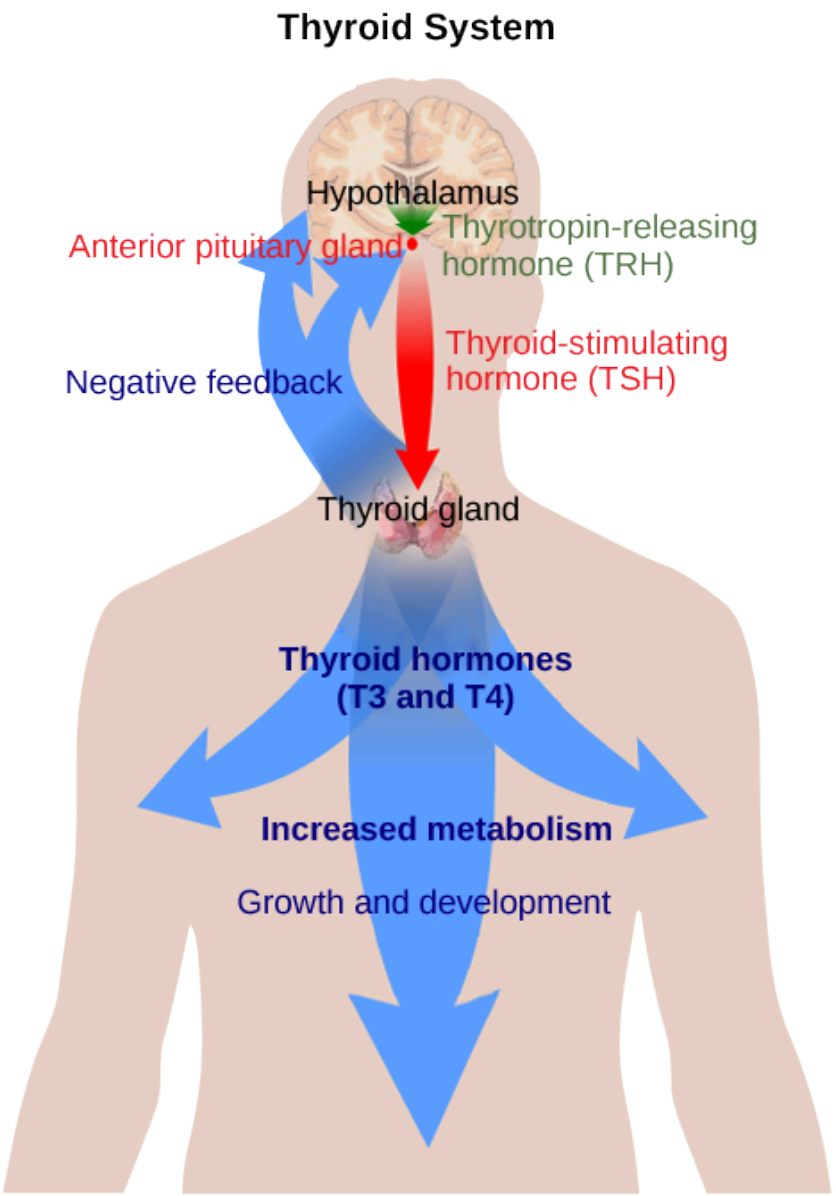


Figure 11.24 The anterior pituitary stimulates the thyroid gland to release thyroid hormones T3 and T4. Increasing levels of these hormones in the blood result in feedback to the hypothalamus and anterior pituitary to inhibit further signaling to the thyroid gland. (credit: modification of work by Mikael Häggström)

Section Summary

Hormones cause cellular changes by binding to receptors on or in target cells. The number of receptors on a target cell can increase or decrease in response to hormone activity.

Hormone levels are primarily controlled through negative feedback, in which rising levels of a hormone inhibit its further release.

The pituitary gland is located at the base of the brain. The anterior pituitary receives signals from the hypothalamus and produces six hormones. The posterior pituitary is an extension of the brain and releases hormones (antidiuretic hormone and oxytocin) produced by the hypothalamus. The thyroid gland is located in the neck and is composed of two lobes. The thyroid produces the hormones thyroxine and triiodothyronine. The thyroid also produces calcitonin. The parathyroid glands lie on the posterior surface of the thyroid gland and produce parathyroid hormone.

The adrenal glands are located on top of the kidneys and consist of the adrenal cortex and adrenal medulla. The adrenal cortex produces the corticosteroids, glucocorticoids and mineralocorticoids. The adrenal medulla is the inner part of the adrenal gland and produces epinephrine and norepinephrine.

The pancreas lies in the abdomen between the stomach and the small intestine. Clusters of endocrine cells in the pancreas form the islets of Langerhans, which contain alpha cells that release glucagon and beta cells that release insulin. Some organs possess endocrine activity as a secondary function but have another primary function. The heart produces the hormone atrial natriuretic peptide, which functions to reduce blood volume, pressure, and Na^+ concentration. The gastrointestinal tract produces various hormones that aid in digestion. The kidneys produce erythropoietin. The thymus produces hormones that aid in the development of the immune system. The gonads produce steroid hormones, including testosterone in males and estrogen and progesterone in females. Adipose tissue produces leptin, which promotes satiety signals in the brain.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4671#h5p-73

Glossary

adrenal gland: the endocrine gland associated with the kidneys
down-regulation: a decrease in the number of hormone receptors in response to increased hormone levels
endocrine gland: the gland that secretes hormones into the surrounding interstitial fluid, which then diffuse into blood and are carried to various organs and tissues within the body
exocrine gland: the gland that secretes chemicals through ducts that lead to skin surfaces, body cavities, and organ cavities.
hormone: a chemical released by cells in one area of the body that affects cells in other parts of the body
intracellular hormone receptor: a hormone receptor in the cytoplasm or nucleus of a cell
pancreas: the organ located between the stomach and the small intestine that contains exocrine and endocrine cells
parathyroid gland: the gland located on the surface of the thyroid that produces parathyroid hormone

pituitary gland: the endocrine gland located at the base of the brain composed of an anterior and posterior region; also called hypophysis
thymus: the gland located behind the sternum that produces thymosin hormones that contribute to the development of the immune system
thyroid gland: an endocrine gland located in the neck that produces thyroid hormones thyroxine and triiodothyronine
up-regulation: an increase in the number of hormone receptors in response to increased hormone levels

11.5 Musculoskeletal System

Learning Objectives

By the end of this section, you will be able to:

- Discuss the axial and appendicular parts of the skeletal system
- Explain the role of joints in skeletal movement
- Explain the role of muscles in locomotion=

The muscular and skeletal systems provide support to the body and allow for movement. The bones of the skeleton protect the body's internal organs and support the weight of the body. The muscles of the muscular system contract and pull on the bones, allowing for movements as diverse as standing, walking, running, and grasping items.

Injury or disease affecting the musculoskeletal system can be very debilitating. The most common musculoskeletal diseases worldwide are caused by malnutrition, which can negatively affect development and maintenance of bones and muscles. Other diseases affect the joints, such as arthritis, which can make movement difficult and, in advanced cases, completely impair mobility.

Progress in the science of prosthesis design has resulted in the development of artificial joints, with joint replacement surgery in the hips and knees being the most common. Replacement joints for shoulders, elbows, and fingers are also available.

Skeletal System

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

Concept in Action

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

The axial skeleton forms the central axis of the body and includes the bones of the skull, ossicles of the middle ear, hyoid bone of the throat, vertebral column, and the thoracic cage (rib cage) (Figure 11.25).

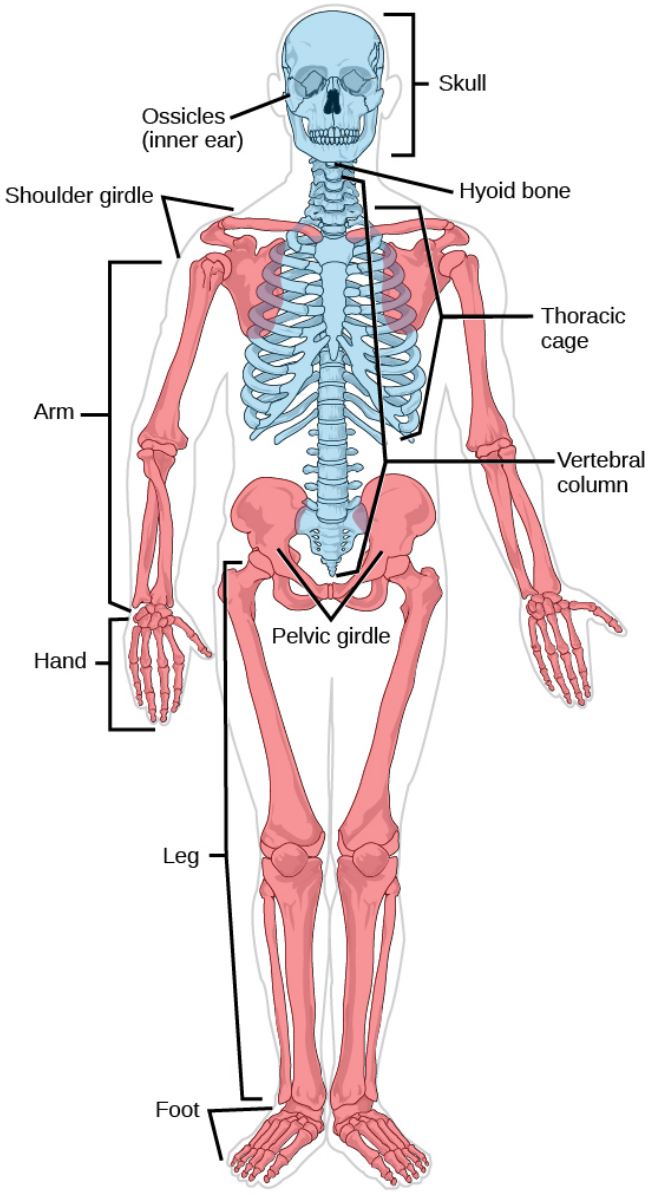


Figure 11.25 The axial skeleton, shown in blue, consists of the bones of the skull, ossicles of the middle ear, hyoid bone, vertebral column, and thoracic cage. The appendicular skeleton, shown in red, consists of the bones of the pectoral limbs, pectoral girdle, pelvic limb, and pelvic girdle. (credit: modification of work by Mariana Ruiz Villareal)

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

The auditory ossicles of the middle ear transmit sounds from the air as vibrations to the fluid-filled cochlea. The auditory ossicles consist of two malleus (hammer) bones, two incus (anvil) bones, and

two stapes (stirrups), one on each side. Facial bones provide cavities for the sense organs (eyes, mouth, and nose), and serve as attachment points for facial muscles.

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

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

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

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

The appendicular skeleton is composed of the bones of the upper and lower limbs. It also includes the pectoral, or shoulder girdle, which attaches the upper limbs to the body, and the pelvic girdle, which attaches the lower limbs to the body (Figure 11.25).

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

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

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

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

Joints and Skeletal Movement

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

There are two ways to classify joints: based on their structure or based on their function. The structural classification divides joints into fibrous, cartilaginous, and synovial joints depending on the material composing the joint and the presence or absence of a cavity in the joint. The bones of fibrous joints are held together by fibrous connective tissue. There is no cavity, or space, present between the bones, so most fibrous joints do not move at all, or are only capable of minor movements. The joints between the bones in the skull and between the teeth and the bone of their sockets are examples of fibrous joints (Figure 11.26 **a**).

Cartilaginous joints are joints in which the bones are connected by cartilage (Figure 11.26 **b**). An example is found at the joints between vertebrae, the so-called “disks” of the backbone. Cartilaginous joints allow for very little movement.

Synovial joints are the only joints that have a space between the adjoining bones (Figure 11.26 **c**). This space is referred to as the joint cavity and is filled with fluid. The fluid lubricates the joint, reducing friction between the bones and allowing for greater movement. The ends of the bones are covered with cartilage and the entire joint is surrounded by a capsule. Synovial joints are capable of the greatest movement of the joint types. Knees, elbows, and shoulders are examples of synovial joints.

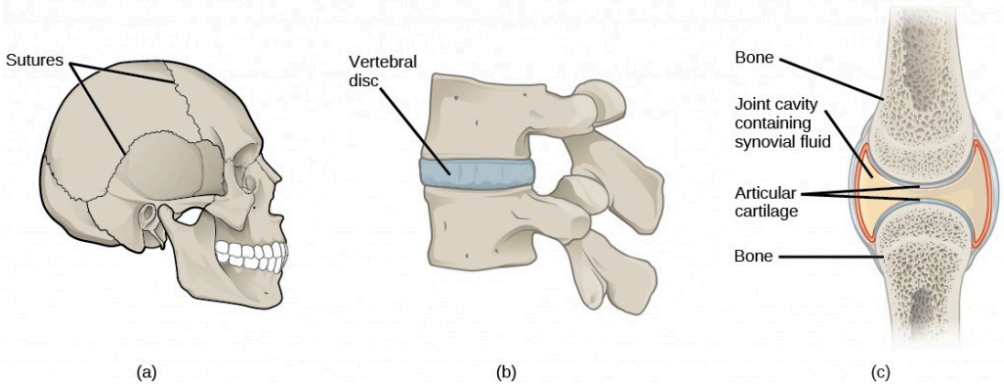


Figure 11.26 (a) Sutures are fibrous joints found only in the skull. (b) Cartilaginous joints are bones connected by cartilage, such as between vertebrae. (c) Synovial joints are the only joints that have a space or “synovial cavity” in the joint.

The wide range of movement allowed by synovial joints produces different types of movements. Angular movements are produced when the angle between the bones of a joint changes. Flexion, or

bending, occurs when the angle between the bones decreases. Moving the forearm upward at the elbow is an example of flexion. Extension is the opposite of flexion in that the angle between the bones of a joint increases. Rotational movement is the movement of a bone as it rotates around its own longitudinal axis. Movement of the head as in saying "no" is an example of rotation.

Rheumatologist

Rheumatologists are medical doctors who specialize in the diagnosis and treatment of disorders of the joints, muscles, and bones. They diagnose and treat diseases such as arthritis, musculoskeletal disorders, osteoporosis, plus autoimmune diseases like ankylosing spondylitis, a chronic spinal inflammatory disease and rheumatoid arthritis.

Rheumatoid arthritis (RA) is an inflammatory disorder that primarily affects synovial joints of the hands, feet, and cervical spine. Affected joints become swollen, stiff, and painful. Although it is known that RA is an autoimmune disease in which the body's immune system mistakenly attacks healthy tissue, the exact cause of RA remains unknown. Immune cells from the blood enter joints and the joint capsule causing cartilage breakdown and swelling of the joint lining. Breakdown of cartilage causes bones to rub against each other causing pain. RA is more common in women than men and the age of onset is usually between 40 to 50 years.

Rheumatologists can diagnose RA based on symptoms such as joint inflammation and pain, x-ray and MRI imaging, and blood tests. Arthrography is a type of medical imaging of joints that uses a contrast agent, such as a dye that is opaque to x-rays. This allows the soft tissue structures of joints—such as cartilage, tendons, and ligaments—to be visualized. An arthrogram differs from a regular x-ray by showing the surface of soft tissues lining the joint in addition to joint bones. An arthrogram allows early degenerative changes in joint cartilage to be detected before bones become affected.

There is currently no cure for RA; however, rheumatologists have a number of treatment options available. Treatments are divided into those that reduce the symptoms of the disease and those that reduce the damage to bone and cartilage caused by the disease. Early stages can be treated with rest of the affected joints through the use of a cane, or with joint splints that minimize inflammation. When inflammation has decreased, exercise can be used to strengthen muscles that surround the joint and to maintain joint flexibility. If joint damage is more extensive, medications can be used to relieve pain and decrease inflammation. Anti-inflammatory drugs that may be used include aspirin, topical pain relievers, and corticosteroid injections. Surgery may be required in cases where joint damage is severe. Physicians are now using drugs that reduce the damage to bones and cartilage caused by the disease to slow its development. These drugs are diverse in their mechanisms but they all act to reduce the impact of the autoimmune response, for example by inhibiting the inflammatory response or reducing the number of T lymphocytes, a cell of the immune system.

Muscles

Muscles allow for movement 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 11.27).

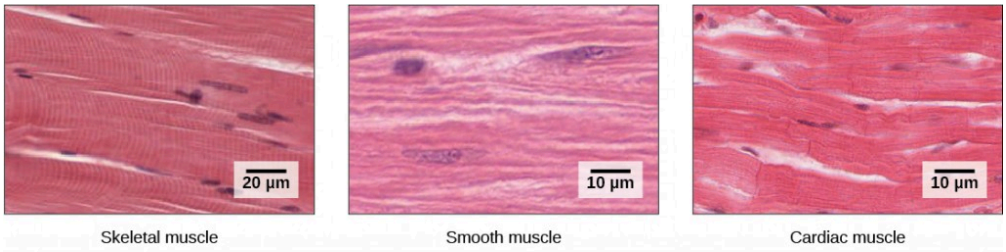


Figure 11.27 The body contains three types of muscle tissue: skeletal muscle, smooth muscle, and cardiac muscle. Notice that skeletal muscle cells are long and cylindrical, they have multiple nuclei, and the small, dark nuclei are pushed to the periphery of the cell. Smooth muscle cells are short, tapered at each end, and have only one nucleus each. Cardiac muscle cells are also cylindrical, but short. The cytoplasm may branch, and they have one or two nuclei 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 and sometimes the skin and control locomotion and any other movement that can be consciously controlled. Because it can be controlled intentionally, skeletal muscle is also called voluntary muscle. When viewed under a microscope, skeletal muscle tissue has a striped or striated appearance. This appearance results from the arrangement of the proteins inside the cell that are responsible for contraction. The cells of skeletal muscle are long and tapered and have multiple nuclei on the periphery of each cell.

Smooth muscle tissue occurs in the walls of hollow organs such as the intestines, stomach, and urinary bladder, and around passages such as in the respiratory tract and blood vessels. Smooth muscle has no striations, is not under voluntary control, and is called involuntary muscle. Smooth muscle cells have a single nucleus.

Cardiac muscle tissue is only found in the heart. The contractions of cardiac muscle tissue 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. The cells of cardiac muscle tissue are connected to each other through intercalated disks and usually have just one nucleus per cell.

Skeletal Muscle Fiber Structure and Function

Each skeletal muscle fiber is a skeletal muscle cell. 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. They attach to the plasma membrane, called the sarcolemma, at their ends, so that as myofibrils shorten, the entire muscle cell contracts (Figure 11.28).

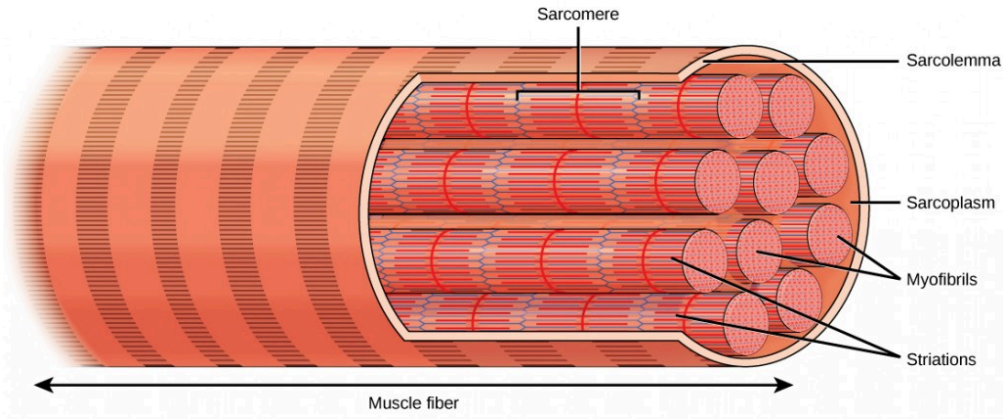


Figure 11.28 A skeletal muscle fiber is surrounded by a plasma membrane called the sarcolemma, with a cytoplasm called the sarcoplasm. A muscle fiber is composed of many fibrils packaged into orderly units. The orderly arrangement of the proteins in each unit, shown as red and blue lines, gives the cell its striated appearance.

The striated appearance of skeletal muscle tissue is a result of repeating bands of the proteins actin and myosin that occur along the length of myofibrils.

Myofibrils are composed of smaller structures called myofilaments. There are two main types of myofilaments: thick filaments and thin filaments. Thick filaments are composed of the protein myosin. The primary component of thin filaments is the protein actin.

The thick and thin filaments alternate with each other in a structure called a sarcomere. The sarcomere is the unit of contraction in a muscle cell. Contraction is stimulated by an electrochemical signal from a nerve cell associated with the muscle fiber. For a muscle cell to contract, the sarcomere must shorten. However, thick and thin filaments do not shorten. Instead, they slide by one another, causing the sarcomere to shorten while the filaments remain the same length. The sliding is accomplished when a molecular extension of myosin, called the myosin head, temporarily binds to an actin filament next to it and through a change in conformation, bends, dragging the two filaments in opposite directions. The myosin head then releases its actin filament, relaxes, and then repeats the process, dragging the two filaments further along each other. The combined activity of many binding sites and repeated movements within the sarcomere causes it to contract. The coordinated contractions of many sarcomeres in a myofibril leads to contraction of the entire muscle cell and ultimately the muscle itself. The movement of the myosin head requires ATP, which provides the energy for the contraction.

Concept in Action

View this animation to see how muscle fibers are organized.

Sliding Filament Model of Contraction

For a muscle cell to contract, the sarcomere must shorten. However, thick and thin filaments—the components of sarcomeres—do not shorten. Instead, they slide by one another, causing the sarcomere to shorten while the filaments remain the same length. The sliding filament theory of muscle contraction was developed to fit the differences observed in the named bands on the sarcomere at different degrees of muscle contraction and relaxation. The mechanism of contraction is the binding of myosin to actin, forming cross-bridges that generate filament movement (Figure 11.29).

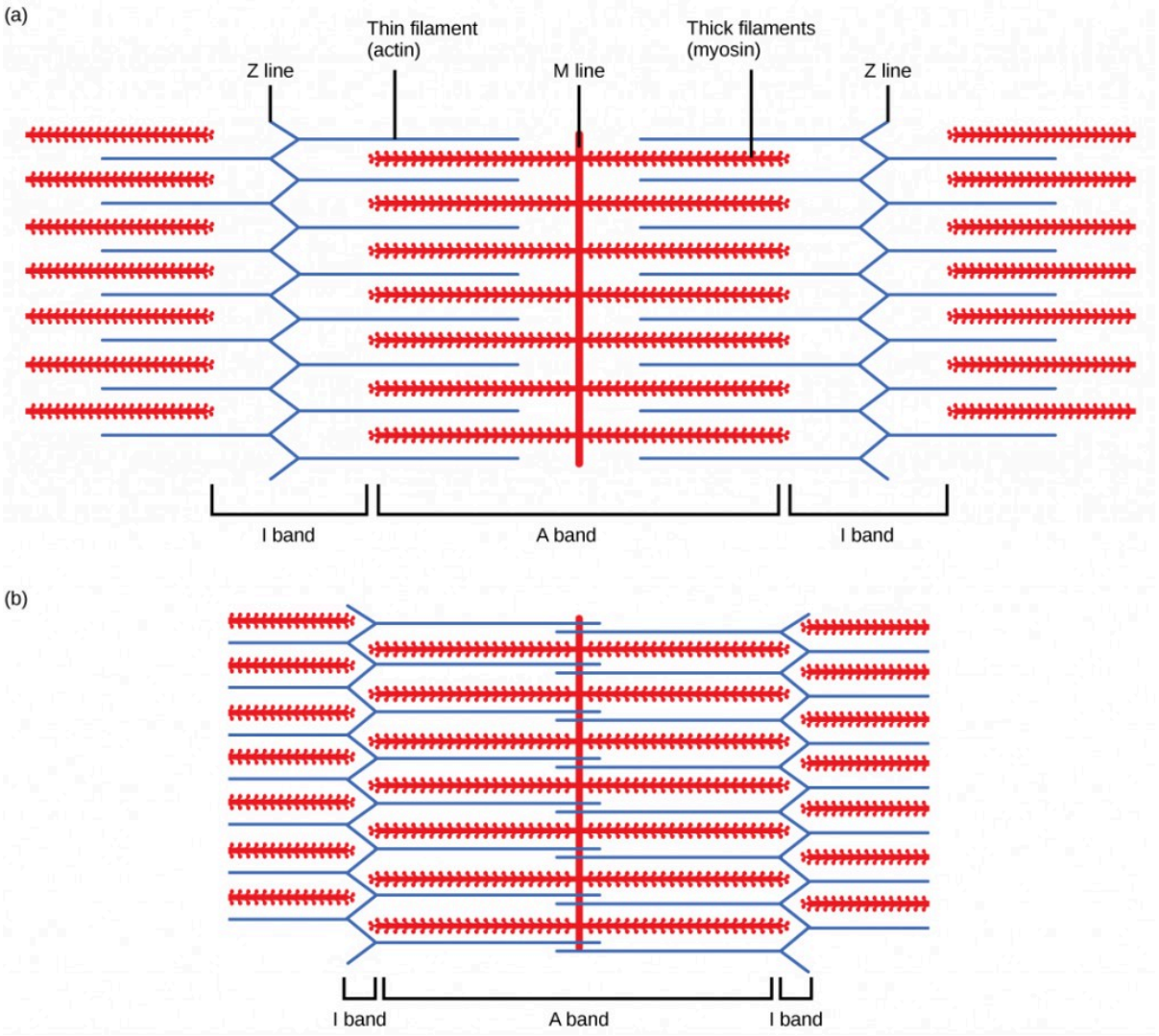


Figure 11.29.
When (a) a sarcomere (b) contracts, the Z lines move closer together and the I band gets smaller. The A band stays the same width and, at full contraction, the thin filaments overlap.

When a sarcomere shortens, some regions shorten whereas others stay the same length. A sarcomere is defined as the distance between two consecutive Z discs or Z lines; when a muscle contracts, the distance between the Z discs is reduced. The H zone—the central region of the A zone—contains only thick filaments and is shortened during contraction. The I band contains only thin filaments and also

shortens. The A band does not shorten—it remains the same length—but A bands of different sarcomeres move closer together during contraction, eventually disappearing. Thin filaments are pulled by the thick filaments toward the center of the sarcomere until the Z discs approach the thick filaments. The zone of overlap, in which thin filaments and thick filaments occupy the same area, increases as the thin filaments move inward.

ATP and Muscle Contraction

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

ATP binding causes myosin to release actin, allowing actin and myosin to detach from each other. After this happens, the newly bound ATP is converted to ADP and inorganic phosphate, 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 11.30).

Concept in Action

Watch this video explaining how a muscle contraction is signaled.

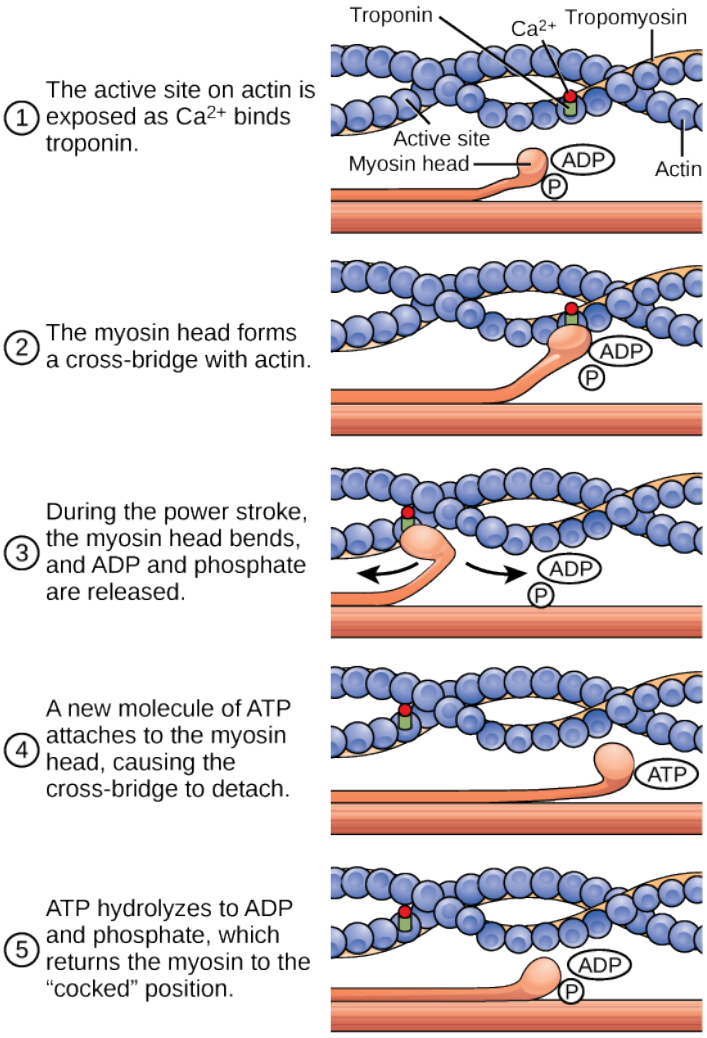


Figure 11.30. The cross-bridge muscle contraction cycle, which is triggered by Ca2+ binding to the actin active site, is shown. With each contraction cycle, actin moves relative to myosin.

Section Summary

The human skeleton is an endoskeleton that is composed of the axial and appendicular skeleton. The axial skeleton is composed of the bones of the skull, ossicles of the ear, hyoid bone, vertebral column, and ribcage. The skull consists of eight cranial bones and 14 facial bones. Six bones make up the ossicles of the middle ear, while the hyoid bone is located in the neck under the mandible. The vertebral column contains 26 bones and surrounds and protects the spinal cord. The thoracic cage consists of the sternum, ribs, thoracic vertebrae, and costal cartilages. The appendicular skeleton is made up of the upper and lower limbs. The pectoral girdle is composed of the clavicles and the scapulae. The upper limb contains 30 bones in the arm, the forearm, and the hand. The pelvic girdle attaches the lower limbs to the axial skeleton. The lower limb includes the bones of the thigh, the leg, and the foot.

The structural classification of joints divides them into fibrous, cartilaginous, and synovial joints. The

bones of fibrous joints are held together by fibrous connective tissue. Cartilaginous joints are joints in which the bones are connected by cartilage. Synovial joints are joints that have a space between the adjoining bones. The movement of synovial joints includes angular and rotational. Angular movements are produced when the angle between the bones of a joint changes. Rotational movement is the movement of a bone as it rotates around its own longitudinal axis.

The body contains three types of muscle tissue: skeletal muscle, cardiac muscle, and smooth muscle. Muscles are composed of individual cells called muscle fibers. Muscle fibers consist of myofilaments composed of the proteins actin and myosin arranged in units called sarcomeres. Contraction of the muscle occurs by the combined action of myosin and actin fibers sliding past each other when the myosin heads bind to the actin fiber, bend, disengage, and then repeat the process.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4682#h5p-74

Glossary

appendicular skeleton: the skeleton composed of the bones of the upper limbs, which function to grasp and manipulate objects, and the lower limbs, which permit locomotion
auditory ossicles: (also, middle ear bones) the bones that transduce sounds from the air into vibrations in the fluid-filled cochlea
axial skeleton: skeleton that forms the central axis of the body and includes the bones of the skull, the ossicles of the middle ear, the hyoid bone of the throat, the vertebral column, and the thoracic cage (ribcage)
cardiac muscle tissue: the muscle tissue found only in the heart; cardiac contractions pump blood throughout the body and maintain blood pressure
cartilaginous joint: a joint in which the bones are connected by cartilage
fibrous joint: a joint held together by fibrous connective tissue
hyoid bone: the bone that lies below the mandible in the front of the neck
joint: the point at which two or more bones meet
myofibril: the long cylindrical structures that lie parallel to the muscle fiber
myofilament: the small structures that make up myofibrils
pectoral girdle: the bones that transmit the force generated by the upper limbs to the axial skeleton
pelvic girdle: the bones that transmit the force generated by the lower limbs to the axial skeleton
sarcolemma: the plasma membrane of a skeletal muscle fiber
sarcomere: the functional unit of skeletal muscle
skeletal muscle tissue: forms skeletal muscles, which attach to bones and control locomotion and any movement that can be consciously controlled

skull: the bone that supports the structures of the face and protects the brain
smooth muscle tissue: the muscle that 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
synovial joints: the only joints that have a space between the adjoining bones
thoracic cage: (also, ribcage) the skeleton of the chest, which consists of the ribs, thoracic vertebrae, sternum, and costal cartilages
vertebral column: (also, spine) the column that surrounds and protects the spinal cord, supports the head, and acts as an attachment point for ribs and muscles of the back and neck

11.6 Nervous System

Learning Objectives

By the end of this section, you will be able to:

- Describe the form and function of a neuron
- Describe the basic parts and functions of the central nervous system
- Describe the basic parts and functions of the peripheral nervous system

As you read this, your nervous system is performing several functions simultaneously. The visual system is processing what is seen on the page; the motor system controls your eye movements and 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. The nervous system is one of two systems that exert control over all the organ systems of the body; the other is the endocrine system. The nervous system's control is much more specific and rapid than the hormonal system. It communicates signals through cells and the tiny gaps between them rather than through the circulatory system as in the endocrine system. It uses a combination of chemical and electrochemical signals, rather than purely chemical signals used by the endocrine system to cover long distances quickly. The nervous system acquires information from sensory organs, processes it and then may initiate a response either through motor function, leading to movement, or in a change in the organism's physiological state.

Nervous systems throughout the animal kingdom vary in structure and complexity. 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." Flatworms have both a central nervous system (CNS), made up of a ganglion (clusters of connected neurons) 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. These ganglia can control movements and behaviors without input from the brain.

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 (toward the stomach) whereas the vertebrate spinal cords are located dorsally (toward the back). 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.

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. There is great diversity in the types of neurons and glia that are present in different parts of the nervous system.

Neurons and Glial Cells

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

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

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 for receiving and sending the electrical signals that make communication between neurons possible (Figure 11.30). 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, most have one or many dendrites.

The bilayer lipid membrane that surrounds a neuron is impermeable to ions. To enter or exit the neuron, ions must pass through ion channels that span the membrane. Some ion channels need to be activated 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. The difference in total charge between the inside and outside of the cell is called the 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). This voltage is called the resting membrane potential; it is caused by differences in the concentrations of ions inside and outside the cell and the selective permeability created by ion channels. Sodium-potassium pumps in the membrane produce the different ion concentrations inside and outside of the cell by bringing in two K^+ ions and removing three Na^+ ions. The actions of this pump are costly: one molecule of ATP is used up for each turn. Up to 50 percent of a neuron's ATP is used in maintaining its membrane resting potential. Potassium ions (K^+), which are higher inside the cell, move fairly freely out of the neuron through potassium channels; this loss of positive charge produces a net negative charge inside the cell. Sodium ions (Na^+), which are low inside, have a driving force to enter but move less freely. Their channels are voltage dependent and will open when a slight change in the membrane potential triggers them.

A neuron can receive input from other neurons and, if this input is strong enough, send the signal to downstream neurons. Transmission of a signal between neurons is generally carried by a chemical, called a neurotransmitter, which diffuses from the axon of one neuron to the dendrite of a second neuron. When neurotransmitter molecules bind to receptors located on a neuron's dendrites, the

neurotransmitter opens ion channels in the dendrite's plasma membrane. This opening allows sodium ions to enter the neuron and results in depolarization of the membrane—a decrease in the voltage across the neuron membrane. Once a signal is received by the dendrite, it then travels passively to the cell body. A large enough signal from neurotransmitters will reach the axon. If it is strong enough (that is, if the threshold of excitation, a depolarization to around –60mV is reached), then depolarization creates a positive feedback loop: as more Na^+ ions enter the cell, the axon becomes further depolarized, opening even more sodium channels at further distances from the cell body. This will cause voltage dependent Na^+ channels further down the axon to open and more positive ions to enter the cell. In the axon, this "signal" will become a self-propagating brief reversal of the resting membrane potential called an action potential.

An action potential is an all-or-nothing event; it either happens or it does not. The threshold of excitation must be reached for the neuron to "fire" an action potential. As sodium ions rush into the cell, depolarization actually reverses the charge across the membrane form -70mv to +30mV. This change in the membrane potential causes voltage-gated K^+ channels to open, and K^+ begins to leave the cell, repolarizing it. At the same time, Na^+ channels inactivate so no more Na^+ enters the cell. K^+ ions continue to leave the cell and the membrane potential returns to the resting potential. At the resting potential, the K^+ channels close and Na^+ channels reset. The depolarization of the membrane proceeds in a wave down the length of the axon. It travels in only one direction because the sodium channels have been inactivated and unavailable until the membrane potential is near the resting potential again; at this point they are reset to closed and can be opened again.

An axon is a tube-like structure that propagates the signal from the cell body to specialized endings called axon terminals. These terminals in turn then synapse with other neurons, muscle, or target organs. When the action potential reaches the axon terminal, this causes the release of neurotransmitter onto the dendrite of another neuron. Neurotransmitters released at axon terminals allow signals to be communicated to these other cells, and the process begins again. Neurons usually have one or two axons, but some neurons do not contain any axons.

Some axons are covered with a special structure called a myelin sheath, which acts as an insulator to keep the electrical signal from dissipating as it travels down the axon. This insulation is important, as the axon from a human motor neuron can be as long as a meter (3.2 ft)—from the base of the spine to the toes. The myelin sheath is produced by glial cells. Along the axon there are periodic gaps in the myelin sheath. These gaps are called nodes of Ranvier and are sites where the signal is "recharged" as it travels along the axon.

It is important to note that a single neuron does not act alone—neuronal communication depends on the connections that neurons make with one another (as well as with other cells, like muscle cells). Dendrites from a single neuron may receive synaptic contact from many other neurons. For example, dendrites from a Purkinje cell in the cerebellum are thought to receive contact from as many as 200,000 other neurons.

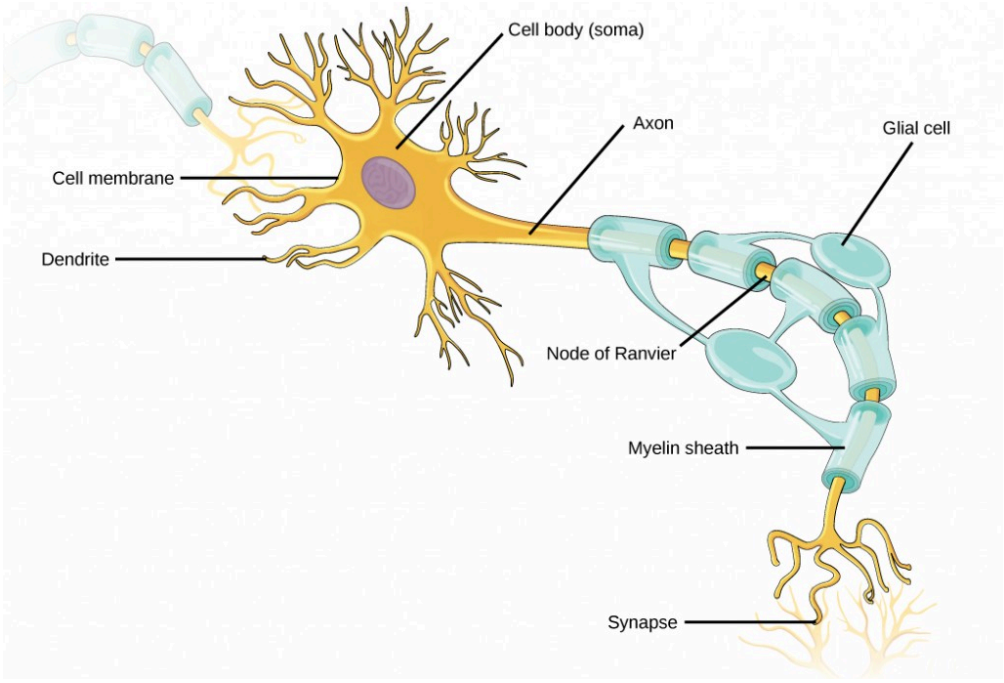


Figure 11.30 Neurons contain organelles common to other cells, such as a nucleus and mitochondria. They also have more specialized structures, including dendrites and axons.

Biology in Action

Neurogenesis

At one time, scientists believed that people were born with all the neurons they would ever have. Research performed during the last few decades indicates that neurogenesis, the birth of new neurons, continues into adulthood. Neurogenesis was first discovered in songbirds that produce new neurons while learning songs. For mammals, new neurons also play an important role in learning: about 1,000 new neurons develop in the hippocampus (a brain structure involved in learning and memory) each day. While most of the new neurons will die, researchers found that an increase in the number of surviving new neurons in the hippocampus correlated with how well rats learned a new task. Interestingly, both exercise and some antidepressant medications also promote neurogenesis in the hippocampus. Stress has the opposite effect. While neurogenesis is quite limited compared to regeneration in other tissues, research in this area may lead to new treatments for disorders such as Alzheimer's, stroke, and epilepsy.

How do scientists identify new neurons? A researcher can inject a compound called bromodeoxyuridine (BrdU) into the brain of an animal. While all cells will be exposed to BrdU, BrdU will only be incorporated into the DNA of newly generated cells that are in S phase. A technique called immunohistochemistry can be used to attach a fluorescent label to the incorporated BrdU, and a researcher can use fluorescent microscopy to visualize the presence of BrdU, and thus new neurons, in brain tissue (Figure 11.31).

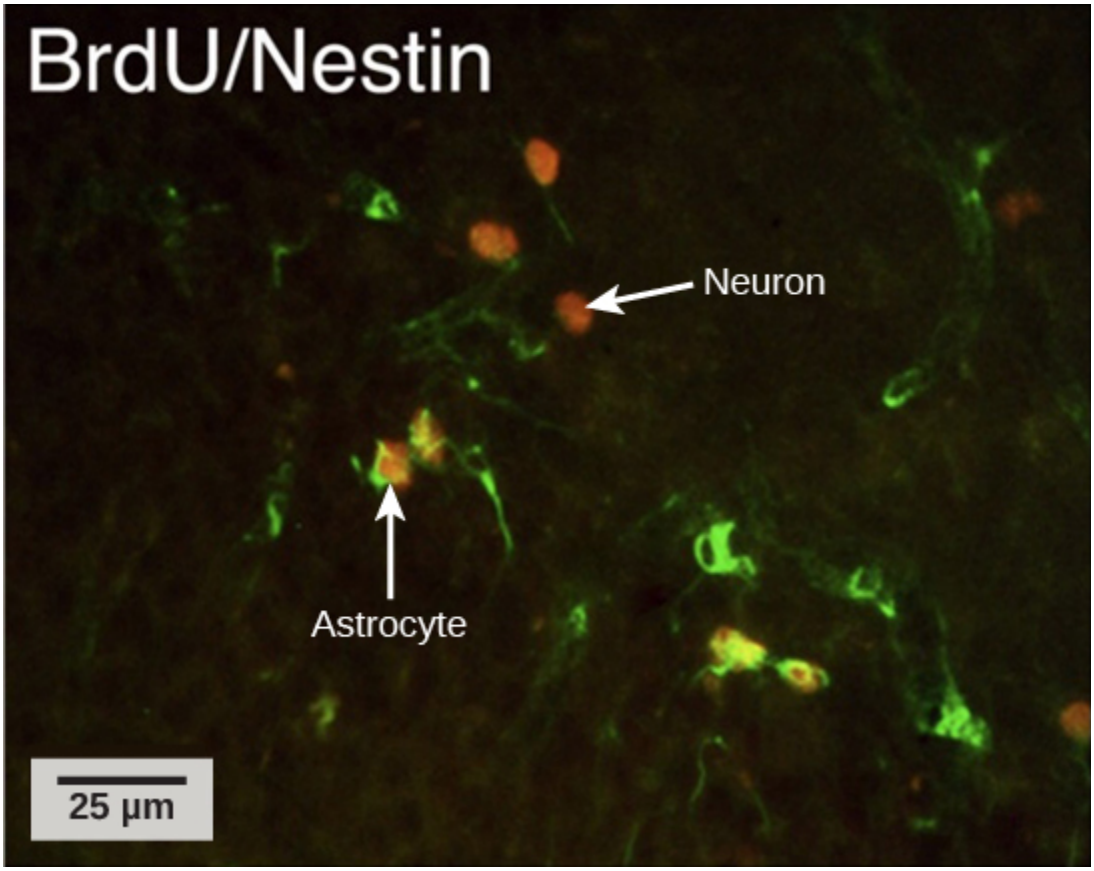


Figure 11.31 This image shows new neurons in a rat hippocampus. New neurons tagged with BrdU glow red in this micrograph. (credit: modification of work by Dr. Maryam Faiz, University of Barcelona)

Concept in Action

Visit this link interactive lab to see more information about neurogenesis, including an interactive laboratory simulation and a video that explains how BrdU labels new cells.

While glial cells are often thought of as the supporting cast of the nervous system, the number of glial cells in the brain actually outnumbers the number of neurons by a factor of 10. Neurons would be unable to function without the vital roles that are fulfilled by these glial cells. Glia guide developing neurons to their destinations, buffer ions and chemicals that would otherwise harm neurons, and provide myelin sheaths around axons. 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. Neurons communicate between the axon of one neuron and the dendrites, and sometimes the cell body, of another neuron across the gap between them, known as the synaptic cleft. When an action potential reaches the end of an axon it stimulates the release of neurotransmitter molecules into the synaptic cleft between the synaptic knob of the axon and the post-synaptic membrane of the dendrite or soma of the next cell. The neurotransmitter is released through exocytosis of vesicles containing the neurotransmitter molecules. The neurotransmitter diffuses across the synaptic cleft and binds to receptors in the post-synaptic membrane. These receptor molecules are chemically regulated ion channels and will open, allowing sodium to enter the cell. If sufficient neurotransmitter has been released an action potential may be initiated in the next cell, but this is not guaranteed. If insufficient neurotransmitter is released the nerve signal will die at this point. There are a number of different neurotransmitters that are specific to neuron types that have specific functions.

The Central Nervous System

The central nervous system (CNS) is made up of the brain and spinal cord and is covered with three layers of protective coverings called meninges ("meninges" is derived from the Greek and means "membranes") (Figure 11.32). The outermost layer is the dura mater, the middle layer is the web-like arachnoid mater, and the inner layer is the pia mater, which directly contacts and covers the brain and spinal cord. The space between the arachnoid and pia maters is filled with cerebrospinal fluid (CSF). The brain floats in CSF, which acts as a cushion and shock absorber.

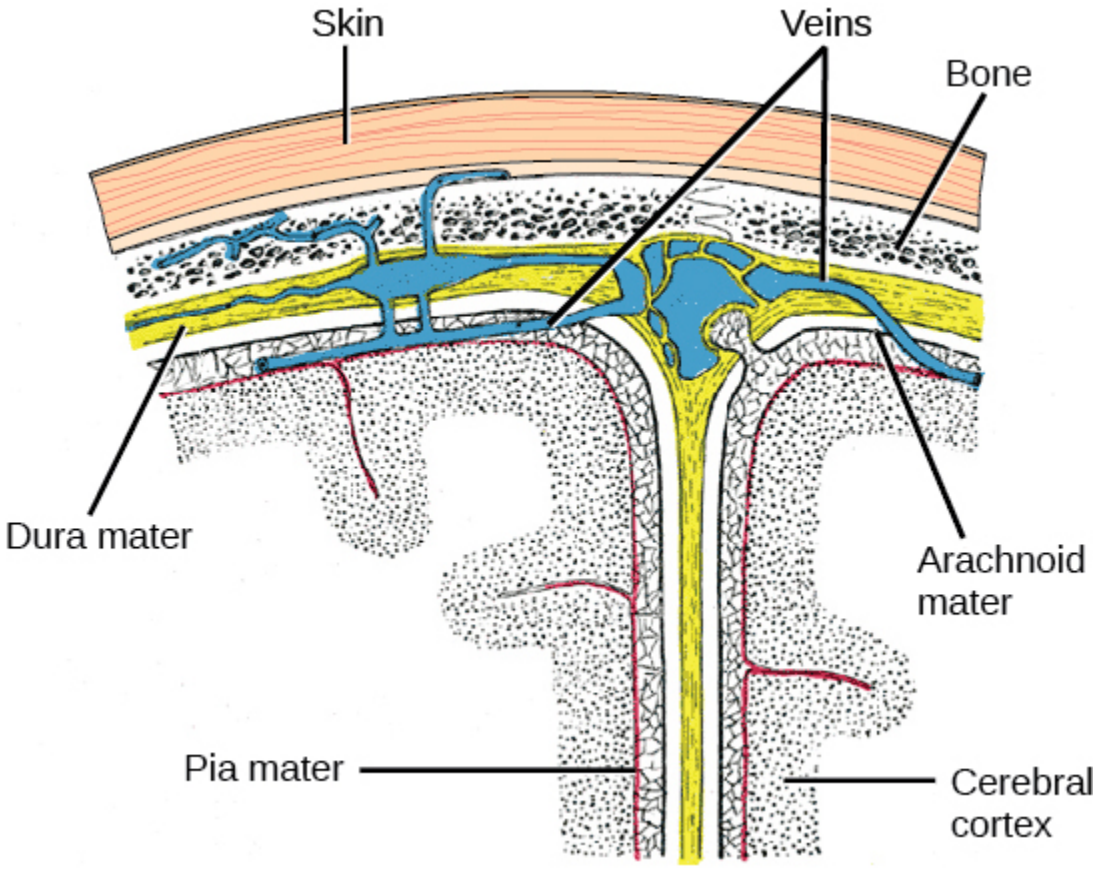


Figure 11.32 The cerebral cortex is covered by three layers of meninges: the dura, arachnoid, and pia maters. (credit: modification of work by Gray's Anatomy)

The 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, cerebellum, brainstem, and retinas. The outermost part of the brain is a thick piece of nervous system tissue called the cerebral cortex. The cerebral cortex, limbic system, and basal ganglia make up the two cerebral hemispheres. A thick fiber bundle called the corpus callosum (corpus = "body"; callosum = "tough") connects the two hemispheres. Although there are some brain functions that are localized more to one hemisphere than the other, the functions of the two hemispheres are largely redundant. In fact, sometimes (very rarely) an entire hemisphere is removed to treat severe epilepsy. While patients do suffer some deficits following the surgery, they can have surprisingly few problems, especially when the surgery is performed on children who have very immature nervous systems.

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

patient is asked to pick up a specific object out of a group of objects with the left hand, the patient will be able to do so but will still be unable to verbally identify it.

Concept in Action

Visit the following website to learn more about split-brain patients and to play a game where you can model split-brain experiments yourself.

Each hemisphere contains regions called lobes that are involved in different functions. Each hemisphere of the mammalian cerebral cortex can be broken down into four functionally and spatially defined lobes: frontal, parietal, temporal, and occipital (Figure 11.33).

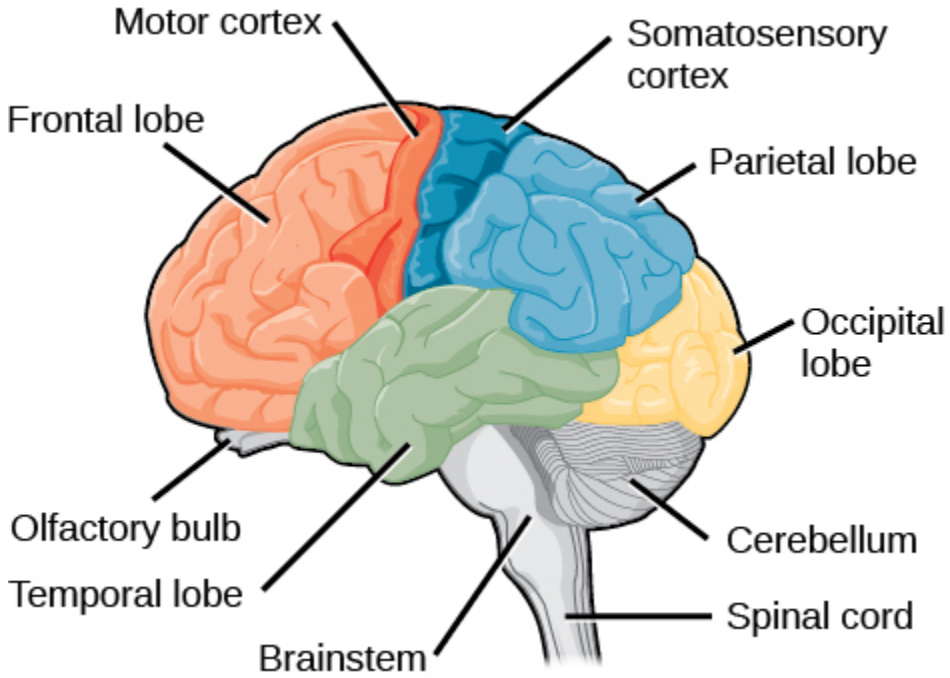


Figure 11.33 The human cerebral cortex includes the frontal, parietal, temporal, and occipital lobes.

The frontal lobe is located at the front of the brain, over the eyes. This lobe contains the olfactory bulb, which processes smells. The frontal lobe also contains the motor cortex, which is important for planning and implementing movement. Areas within the motor cortex map to different muscle groups. 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. 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 and is primarily involved in processing and interpreting sounds. It also contains the hippocampus (named from the Greek for "seahorse," which it resembles in shape) a structure that processes memory formation. 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).

Interconnected brain areas called the basal ganglia play important roles in movement control and posture. The basal ganglia also regulate motivation.

The thalamus acts as a gateway to and from the cortex. It receives sensory and motor inputs from the body and also receives feedback from the cortex. This feedback mechanism can modulate conscious awareness of sensory and motor inputs depending on the attention and arousal state of the animal. The thalamus helps regulate consciousness, arousal, and sleep states.

Below the thalamus is the hypothalamus. The hypothalamus controls the endocrine system by sending signals to the pituitary gland. Among other functions, the hypothalamus is the body's thermostat—it makes sure the body temperature is kept at appropriate levels. Neurons within the hypothalamus also regulate circadian rhythms, sometimes called sleep cycles.

The limbic system is a connected set of structures that regulates emotion, as well as behaviors related to fear and motivation. It plays a role in memory formation and includes parts of the thalamus and hypothalamus as well as the hippocampus. One important structure within the limbic system is a temporal lobe structure called the amygdala. The two amygdala (one on each side) are important both for the sensation of fear and for recognizing fearful faces.

The cerebellum (cerebellum = "little brain") 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. The cerebellum of birds is large compared to other vertebrates because of the coordination required by flight.

The brainstem connects the rest of the brain with the spinal cord and regulates some of the most important and basic functions of the nervous system including breathing, swallowing, digestion, sleeping, 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. 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 meninges and the bones of the vertebral 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 (Figure 11.34). Axons make up the "white matter" and neuron and glia cell bodies (and interneurons) make up the "gray matter." Axons and cell bodies in the dorsa spinal cord convey mostly sensory information from the body to the brain. Axons and cell bodies in the 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).

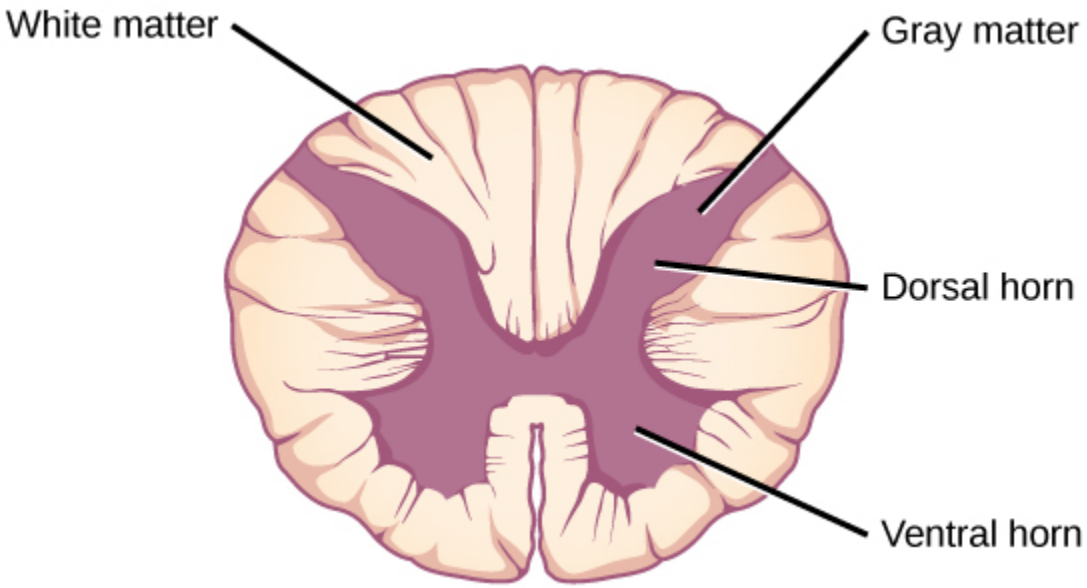


Figure 11.34 A cross-section of the spinal cord shows gray matter (containing cell bodies and interneurons) and white matter (containing myelinated axons).

The Peripheral Nervous System

The peripheral nervous system (PNS) is the connection between the central nervous system and the rest of the body. The PNS can be broken down into the autonomic nervous system, which controls bodily functions without conscious control, and the sensory-somatic nervous system, which transmits sensory information from the skin, muscles, and sensory organs to the CNS and sends motor commands from the CNS to the muscles.

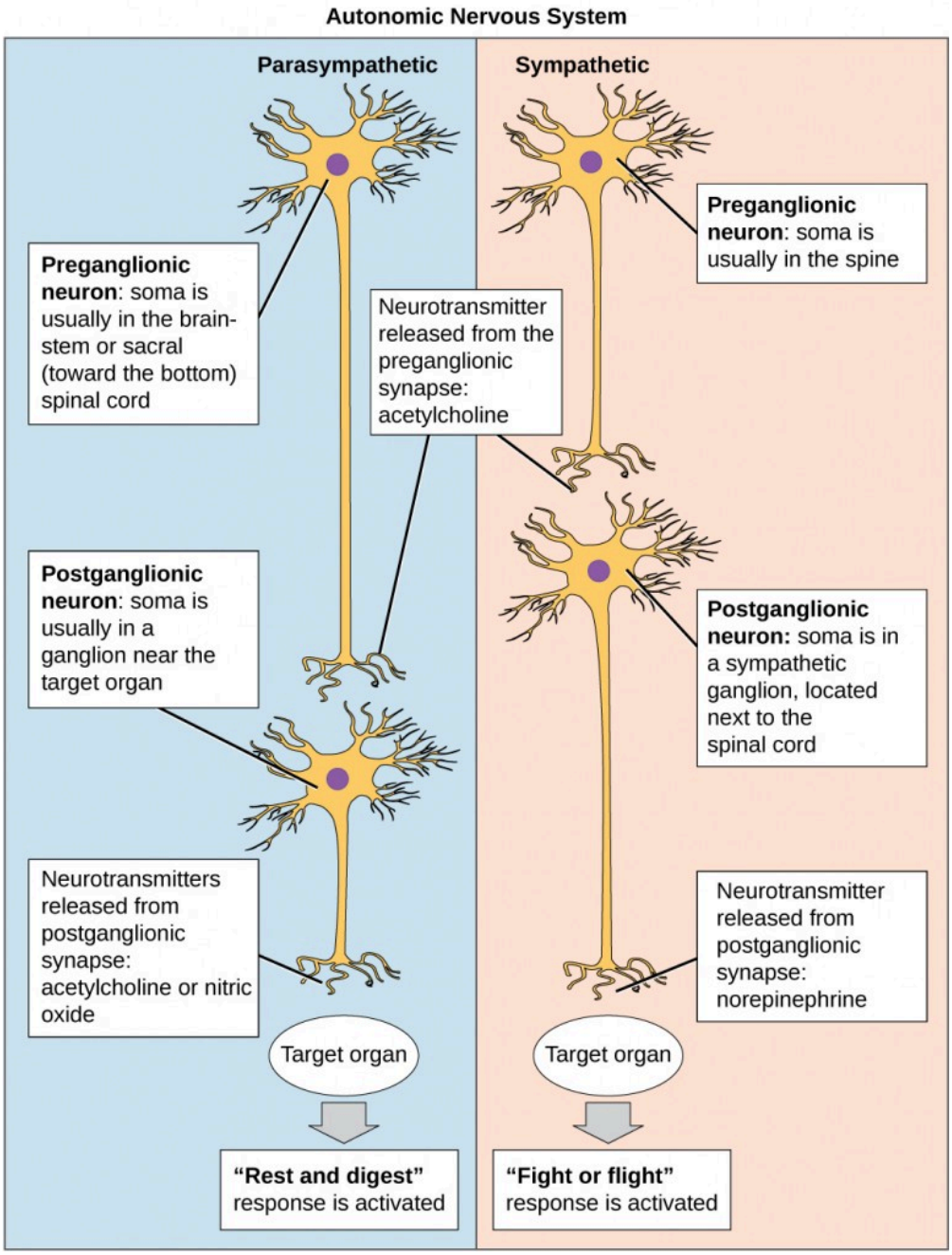


Figure 11.35 In the autonomic nervous system, a preganglionic neuron (originating in the CNS) synapses to a neuron in a ganglion that, in turn, synapses on a target organ. Activation of the sympathetic nervous system causes release of norepinephrine on the target organ. Activation of the parasympathetic nervous system causes release of acetylcholine on the target organ.

The autonomic nervous system serves as the relay between the CNS and the internal organs. It controls the lungs, the heart, smooth muscle, and exocrine and endocrine glands. The autonomic nervous system controls these organs largely without conscious control; it can continuously monitor the conditions of these different systems and implement changes as needed. Signaling to the target tissue usually involves two synapses: a preganglionic neuron (originating in the CNS) synapses to a neuron in a ganglion that, in turn, synapses on the target organ (Figure 11.35). There are two divisions of the

autonomic nervous system that often have opposing effects: the sympathetic nervous system and the parasympathetic nervous system.

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

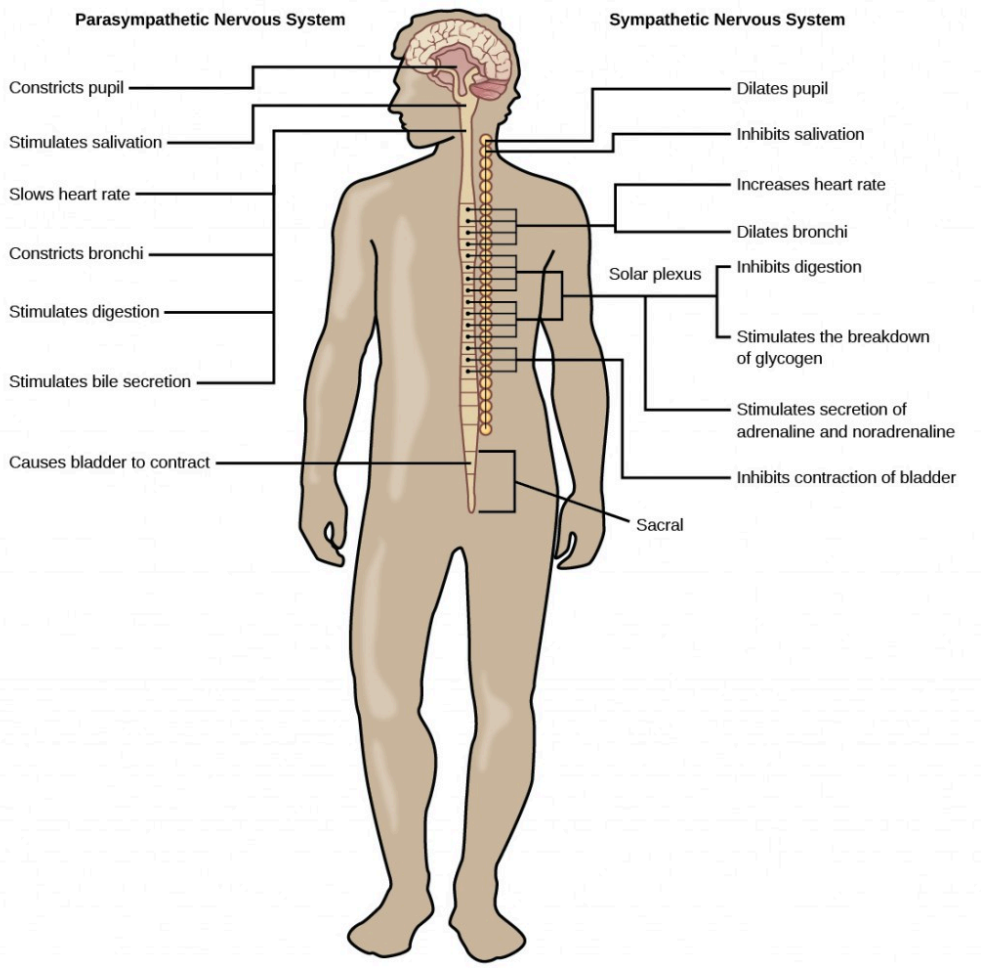


Figure 11.36 The sympathetic and parasympathetic nervous systems often have opposing effects on target organs.

While the sympathetic nervous system is activated in stressful situations, the parasympathetic nervous system allows an animal to "rest and digest." One way to remember this is to think that during a restful situation like a picnic, the parasympathetic nervous system is in control ("picnic" and "parasympathetic" both start with "p"). Parasympathetic preganglionic neurons have cell bodies located in the brainstem and in the sacral (toward the bottom) spinal cord (Figure 11.36). The parasympathetic nervous system resets organ function after the sympathetic nervous system is activated including slowing of heart rate, lowered blood pressure, and stimulation of digestion.

The sensory-somatic nervous system is made up of cranial and spinal nerves and contains both sensory

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

Section Summary

The nervous system is made up of neurons and glia. Neurons are specialized cells that are capable of sending electrical as well as chemical signals. Most neurons contain dendrites, which receive these signals, and axons that send signals to other neurons or tissues. Glia are non-neuronal cells in the nervous system that support neuronal development and signaling. There are several types of glia that serve different functions.

Neurons have a resting potential across their membranes and when they are stimulated by a strong enough signal from another neuron an action potential may carry an electrochemical signal along the neuron to a synapse with another neuron. Neurotransmitters carry signals across synapses to initiate a response in another neuron.

The vertebrate central nervous system contains the brain and the spinal cord, which are covered and protected by three meninges. The brain contains structurally and functionally defined regions. In mammals, these include the cortex (which can be broken down into four primary functional lobes: frontal, temporal, occipital, and parietal), basal ganglia, thalamus, hypothalamus, limbic system, cerebellum, and brainstem—although structures in some of these designations overlap. While functions may be primarily localized to one structure in the brain, most complex functions, like language and sleep, involve neurons in multiple brain regions. The spinal cord is the information superhighway that connects the brain with the rest of the body through its connections with peripheral nerves. It transmits sensory and motor input and also controls motor reflexes.

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

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4693#h5p-75

Glossary

action potential: a momentary change in the electrical potential of a neuron (or muscle) membrane
amygdala: a structure within the limbic system that processes fear
autonomic nervous system: the part of the peripheral nervous system that controls bodily functions
axon: a tube-like structure that propagates a signal from a neuron's cell body to axon terminals
basal ganglia: an interconnected collections of cells in the brain that are involved in movement and motivation
brainstem: a portion of brain that connects with the spinal cord; controls basic nervous system functions like breathing and swallowing
central nervous system (CNS): the nervous system made up of the brain and spinal cord; covered with three layers of protective meninges
cerebellum: the brain structure involved in posture, motor coordination, and learning new motor actions
cerebral cortex: the outermost sheet of brain tissue; involved in many higher-order functions
cerebrospinal fluid (CSF): a clear liquid that surrounds the brain and fills its ventricles and acts as a shock absorber
corpus callosum: a thick nerve bundle that connects the cerebral hemispheres
dendrite: a structure that extends away from the cell body to receive messages from other neurons
depolarization: a change in the membrane potential to a less negative value
frontal lobe: the part of the cerebral cortex that contains the motor cortex and areas involved in planning, attention, and language
glia: (also, glial cells) the cells that provide support functions for neurons
hippocampus: the brain structure in the temporal lobe involved in processing memories
hypothalamus: the brain structure that controls hormone release and body homeostasis
limbic system: a connected brain area that processes emotion and motivation
membrane potential: a difference in electrical potential between the inside and outside of a cell
meninges: (singular: meninx) the membranes that cover and protect the central nervous system
myelin sheath: a cellular extension containing a fatty substance produced by glia that surrounds and insulates axons
neuron: a specialized cell that can receive and transmit electrical and chemical signals
occipital lobe: the part of the cerebral cortex that contains visual cortex and processes visual stimuli
parasympathetic nervous system: the division of autonomic nervous system that regulates visceral functions during relaxation
parietal lobe: the part of the cerebral cortex involved in processing touch and the sense of the body in space
peripheral nervous system (PNS): the nervous system that serves as the connection between the central

nervous system and the rest of the body; consists of the autonomic nervous system and the sensory-somatic nervous system
sensory-somatic nervous system: the system of sensory and motor nerves
spinal cord: a thick fiber bundle that connects the brain with peripheral nerves; transmits sensory and motor information; contains neurons that control motor reflexes
sympathetic nervous system: the division of autonomic nervous system activated during stressful "fight-or-flight" situations
synapse: a junction between two neurons where neuronal signals are communicated
synaptic cleft: a space between the presynaptic and postsynaptic membranes
temporal lobe: the part of the cerebral cortex that processes auditory input; parts of the temporal lobe are involved in speech, memory, and emotion processing
thalamus: the brain area that relays sensory information to the cortex
threshold of excitation: the level of depolarization needed for an action potential to fire

Chapter 12: Introduction to the Immune System and Disease

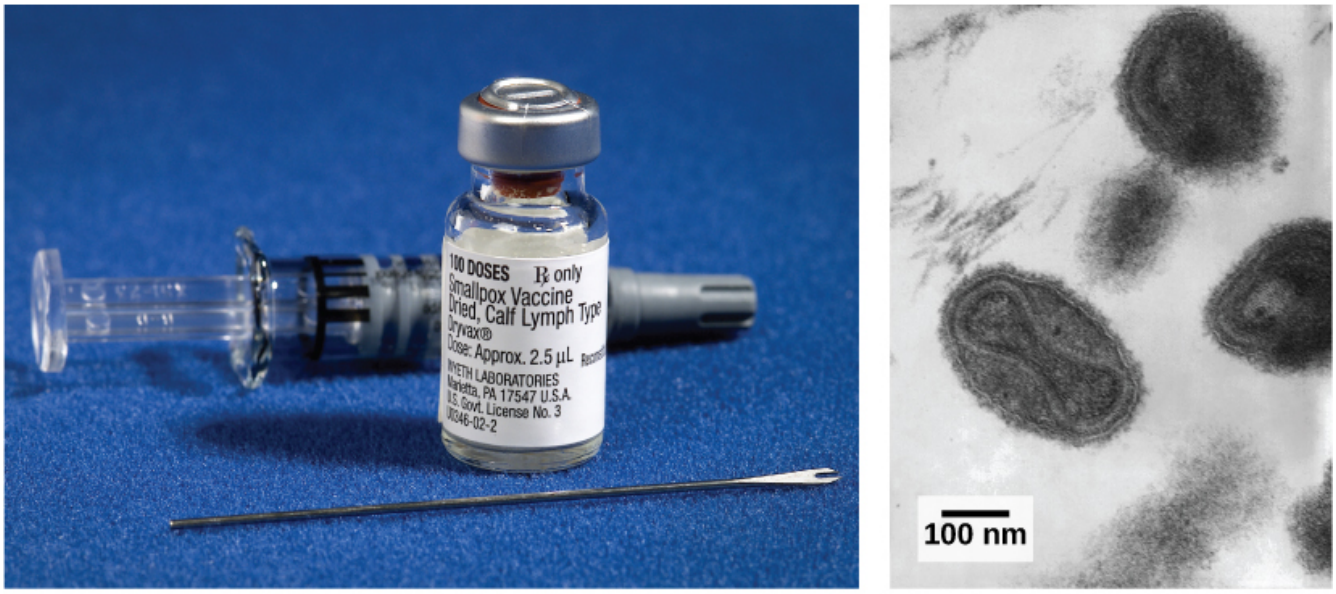


Figure 12.1 (a) This smallpox (variola) vaccine is derived from calves exposed to cowpox virus. Vaccines provoke a reaction in the immune system that prepares it for a subsequent infection by smallpox. (b) Viewed under a transmission electron microscope, you can see the variola's dumbbell-shaped structure that contains the viral DNA. (credit a: modification of work by James Gathany, CDC; credit b: modification of work by Dr. Fred Murphy; Sylvia Whitfield, CDC; scale-bar data from Matt Russell)

Organisms have a wide array of adaptations for preventing attacks of parasites and diseases. The vertebrate defense systems, including those of humans, are complex and multilayered, with defenses unique to vertebrates. These unique vertebrate defenses interact with other defense systems inherited from ancestral lineages, and include complex and specific pathogen recognition and memory mechanisms. Research continues to unravel the complexities and vulnerabilities of the immune system.

Despite a poor understanding of the workings of the body in the early 18th century in Europe, the practice of inoculation as a method to prevent the often-deadly effects of smallpox was introduced from the courts of the Ottoman Empire. The method involved causing limited infection with the smallpox virus by introducing the pus of an affected individual to a scratch in an uninfected person. The resulting infection was milder than if it had been caught naturally and mortality rates were shown to be about two percent rather than 30 percent from natural infections. Moreover, the inoculation gave the individual immunity to the disease. It was from these early experiences with inoculation that the methods of vaccination were developed, in which a weakened or relatively harmless (killed) derivative

of a pathogen is introduced into the individual. The vaccination induces immunity to the disease with few of the risks of being infected. A modern understanding of the causes of the infectious disease and the mechanisms of the immune system began in the late 19th century and continues to grow today.

Search for Key Points in Chapter 12

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4696#h5p-76

12.1 Viruses

Learning Objectives

By the end of this section, you will be able to:

- Describe how viruses were first discovered and how they are detected
- Explain the detailed steps of viral replication
- Describe how vaccines are used in prevention and treatment of viral diseases

Figure 12.2 (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 historical footprints such as 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 domain because they are not considered alive. 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. Living things grow, metabolize, and reproduce. Viruses replicate, but to do so, they are entirely dependent on their host cells. They do not metabolize or grow, but are assembled in their mature form.

Viruses are diverse. They vary in their structure, their replication methods, and in their target hosts or even host cells. While most biological diversity can be understood through evolutionary history, such as how species have adapted to conditions and environments, much about virus origins and evolution remains unknown.

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). These 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 viruses with a light microscope, with the exception of some large virions of the poxvirus family (Figure 12.3).

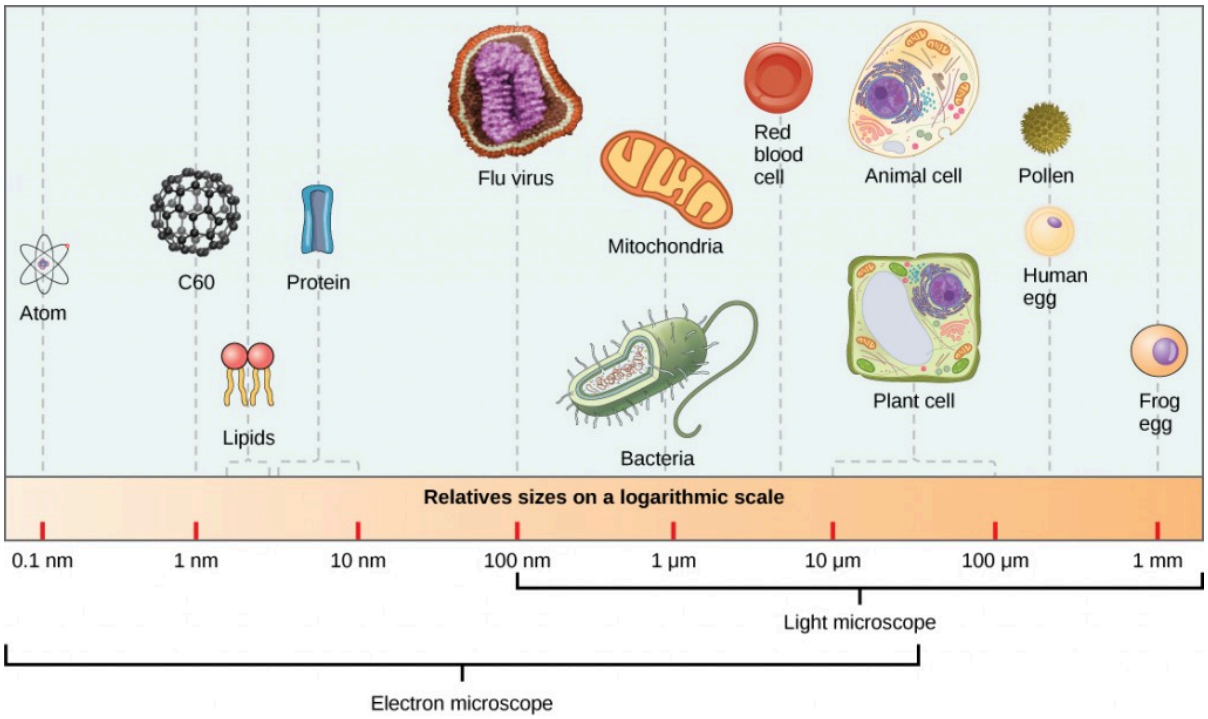


Figure 12.3 The size of a virus is very small relative to the size of cells and organelles.

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 12.2) 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 12.4).

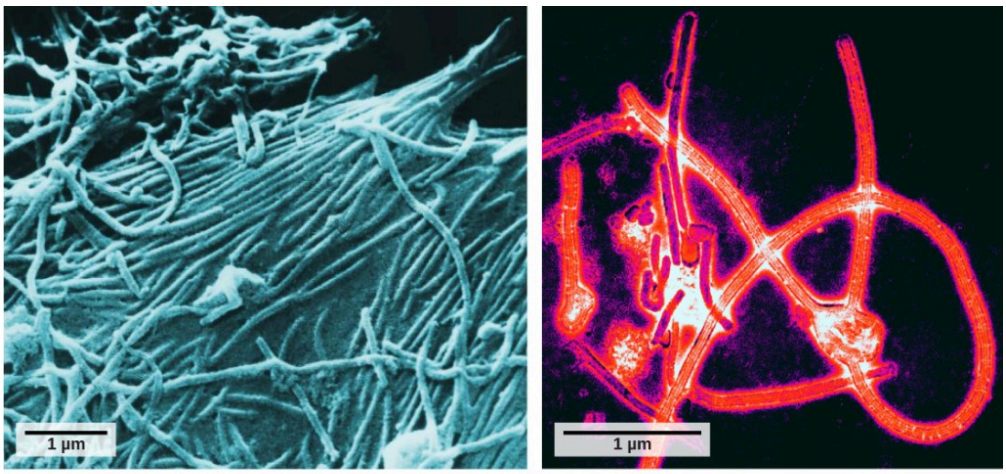


Figure 12.4 The ebola virus is shown here as visualized through (a) a scanning electron micrograph and (b) a transmission electron micrograph. (credit a: modification of work by Cynthia Goldsmith, CDC; credit b: modification of work by Thomas W. Geisbert, Boston University School of Medicine; scale-bar 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.

A virion consists of a nucleic-acid core, an outer protein coating, and sometimes an outer envelope made of protein and phospholipid membranes 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 12.5). 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. For example, the measles virus uses a cell-surface glycoprotein in humans that normally functions in immune reactions and possibly in the sperm-egg interaction at fertilization. Attachment is a requirement for viruses to later penetrate the cell membrane, inject the viral genome, and complete their replication inside the cell.

The T4 bacteriophage, which infects the *E. coli* bacterium, is among the most complex virion 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 12.5). 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.

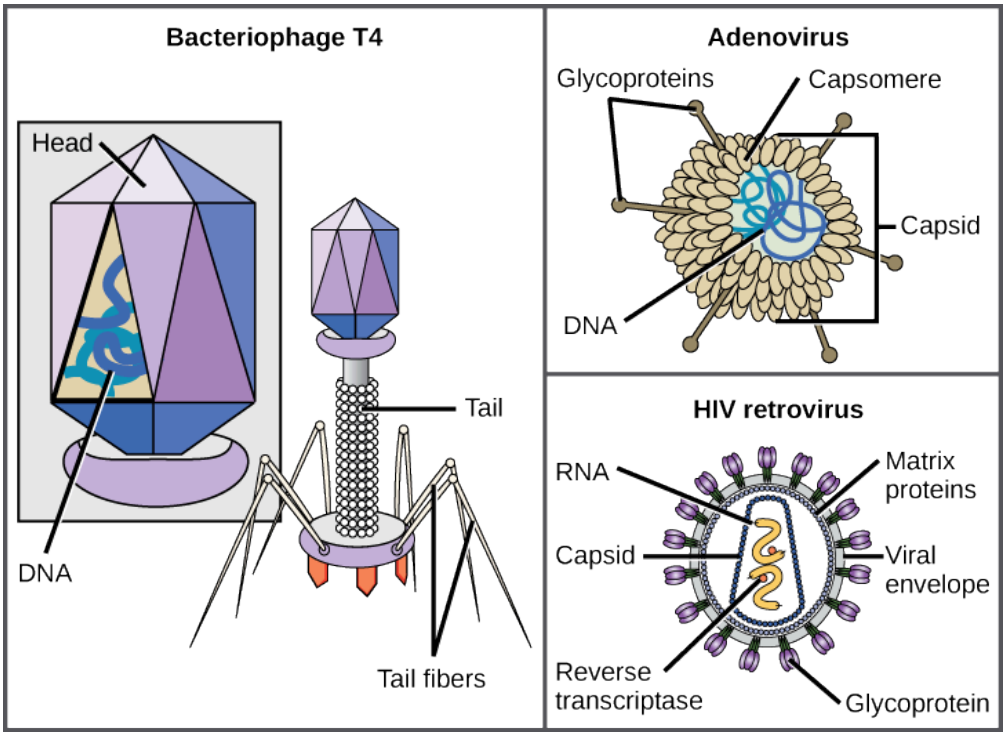


Figure 12.5 Viruses can be complex in shape or relatively simple. This figure shows three relatively complex virions: the bacteriophage T4, with its DNA-containing head group and tail fibers that attach to host cells; adenovirus, which uses spikes from its capsid to bind to the host cells; and HIV, which 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)

Which of the following statements about virus structure is true?

A) All viruses are encased in a viral membrane.

B) The capsomere is made up of small protein subunits called capsids.

C) DNA is the genetic material in all viruses.

D) Glycoproteins help the virus attach to the host cell.

<!–D–>

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.

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. Human diseases caused by RNA viruses include hepatitis C, measles, and rabies.

Viruses can be seen as obligate intracellular parasites. The virus must attach to a living cell, be taken inside, manufacture its proteins and copy its genome, and find a way to escape the cell so the virus can infect other cells and ultimately other individuals. Viruses can infect only certain species of hosts and only certain cells within that host. The molecular basis for this specificity is that a particular surface molecule, known as the viral receptor, must be found on the host cell surface for the virus to attach. Also, metabolic differences seen in different cell types based on differential gene expression are a likely factor in which cells a virus may use to replicate. The cell must be making the substances the virus needs, such as enzymes the virus genome itself does not have genes for, or the virus will not be able to replicate using that cell.

Steps of Virus Infections

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

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-cell 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.

Concept in Action

View this video for a visual explanation of how HIV and influenza attack the body.

Unlike animal viruses, the nucleic acid of bacteriophages is injected into the host cell naked, leaving the capsid outside the cell. Plant and animal viruses can enter their cells through endocytosis, in which the cell membrane surrounds and engulfs the entire virus. Some enveloped viruses enter the cell when the viral envelope fuses directly with the cell 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.

The replication mechanism depends on the viral genome. DNA viruses usually use host cell proteins and enzymes to make additional DNA that is used to copy the genome or be transcribed to messenger RNA (mRNA), which is then used in protein synthesis. RNA viruses, such as the influenza virus, usually use the RNA core 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 12.6). Of course, there are exceptions to this pattern. If a host cell does not provide the enzymes necessary for viral replication, viral genes supply the information to direct synthesis of the missing proteins. Retroviruses, such as HIV, have an RNA genome that must be reverse transcribed to make DNA, which then is inserted into the host's DNA. To convert RNA into DNA, retroviruses contain genes that encode the virus-specific enzyme reverse transcriptase that transcribes an RNA template to DNA. The fact that HIV produces some of its own enzymes, which are not found in the host, has allowed researchers to develop drugs that inhibit these enzymes. These drugs, including the reverse transcriptase inhibitor

AZT, inhibit HIV replication by reducing the activity of the enzyme without affecting the host's metabolism.

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.

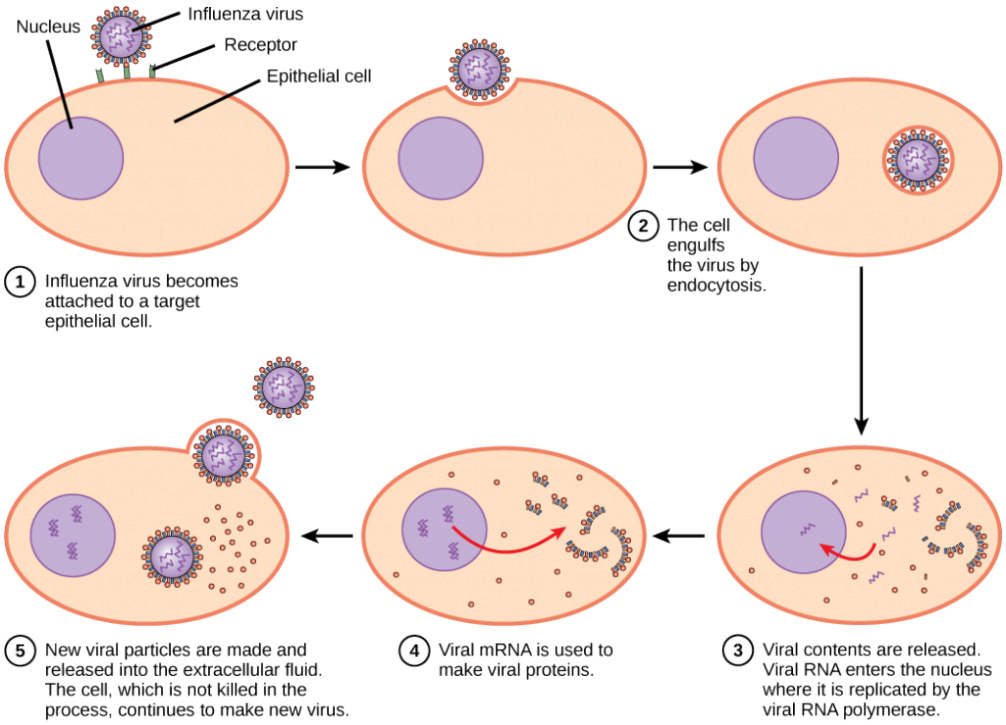


Figure 12.6 In influenza virus infection, glycoproteins attach to a host epithelial cell. As a result, the virus is engulfed. RNA and proteins are made and assembled into new virions.

Influenza virus is packaged in a viral envelope, which fuses with the plasma membrane. This way, the virus can exit the host cell without killing it. What advantage does the virus gain by keeping the host cell alive?

<!–The host cell can continue to make new virus particles.–>

Concept in Action

Click through this tutorial on viruses to identify structures, modes of transmission, replication, and more.

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 12.7). 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.

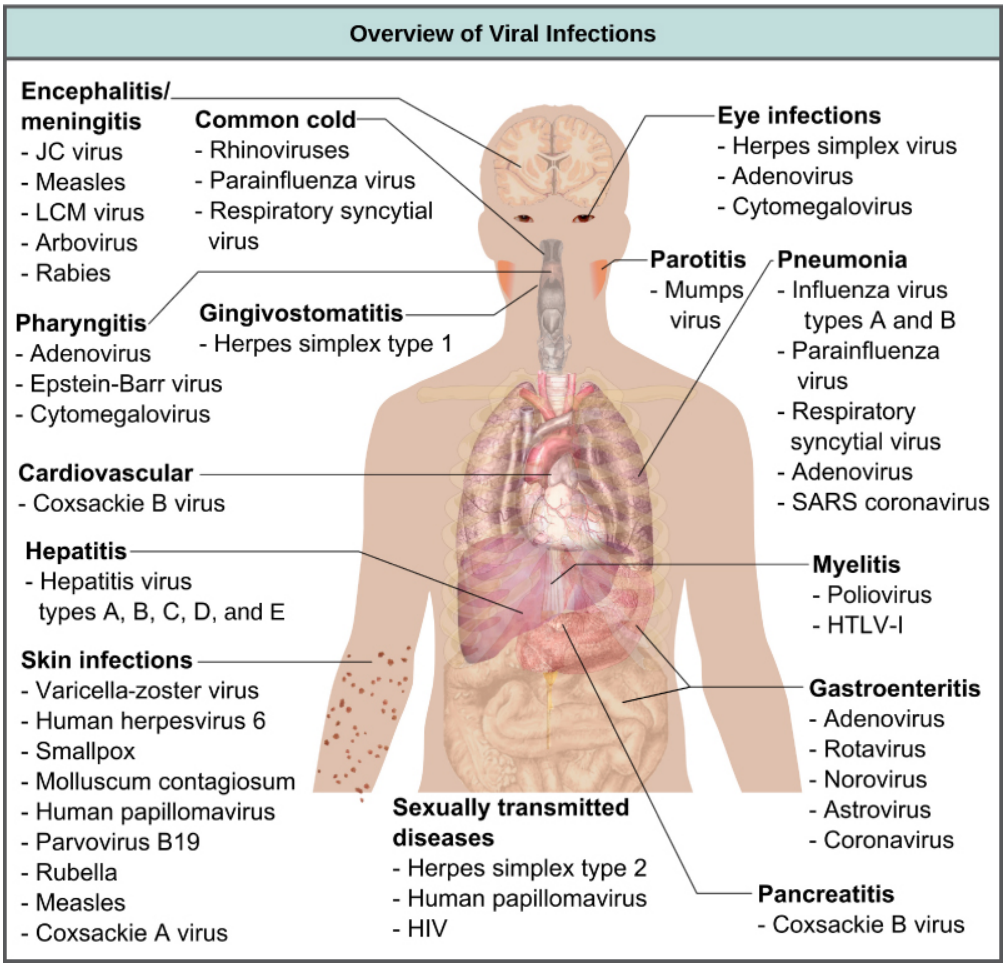


Figure 12.7 Viruses are the cause of dozens of ailments in humans, ranging from mild illnesses to serious diseases. (credit: modification of work by Mikael Häggström)

Vaccines for Prevention

While we do have limited numbers of effective antiviral drugs, such as those used to treat HIV and influenza, the primary method of controlling viral disease is by vaccination, which is intended to prevent outbreaks by building immunity to a virus or virus family. A vaccine may be prepared using weakened live viruses, killed viruses, or molecular subunits of the virus. In general, live viruses lead to better immunity, but have the possibility of causing disease at some low frequency. Killed viral vaccine and the subunit viruses are both incapable of causing disease, but in general lead to less effective or long-lasting immunity.

Weakened live viral vaccines are designed in the laboratory to cause few symptoms in recipients while giving them immunity against future infections. Polio was one disease that represented a milestone in the use of vaccines. Mass immunization campaigns in the U.S. in the 1950s (killed vaccine) and 1960s (live vaccine) essentially eradicated the disease, which caused muscle paralysis in children and generated fear in the general population when regional epidemics occurred. The success of the polio vaccine paved the way for the routine dispensation of childhood vaccines against measles, mumps, rubella, chickenpox, and other diseases.

Live vaccines are usually made by attenuation (weakening) of the "wild-type" (disease-causing) virus by growing it in the laboratory in tissues or at temperatures different from what the virus is accustomed to in the host. For example, the virus may be grown in cells in a test tube, in bird embryos, or in live animals. The adaptation to these new cells or temperature induces mutations in the virus' genomes, allowing them to grow better in the laboratory while inhibiting their ability to cause disease when reintroduced into the conditions found in the host. These attenuated viruses thus still cause an infection, but they do not grow very well, allowing the immune response to develop in time to prevent major disease. The danger of using live vaccines, which are usually more effective than killed vaccines, is the low but significant risk that these viruses will revert back to their disease-causing form by back mutations. Back mutations occur when the vaccine undergoes mutations in the host such that it readapts to the host and can again cause disease, which can then be spread to other humans in an epidemic. This happened as recently as 2007 in Nigeria where mutations in a polio vaccine led to an epidemic of polio in that country.

Some vaccines are in continuous development because certain viruses, such as influenza and HIV, have a high mutation rate compared to other viruses or host cells. With influenza, mutation in genes for the surface molecules helps the virus evade the protective immunity that may have been obtained in a previous influenza season, making it necessary for individuals to get vaccinated every year. Other viruses, such as those that cause the childhood diseases measles, mumps, and rubella, mutate so little that the same vaccine is used year after year.

Vaccines and Antiviral Drugs for Treatment

In some cases, vaccines can be used to treat an active viral infection. In the case of rabies, a fatal neurological disease transmitted in the saliva of rabies virus-infected animals, the progression of the disease from the time of the animal bite to the time it enters the central nervous system may be two weeks or longer. This is enough time to vaccinate an individual who suspects being bitten by a rabid animal, and the boosted immune response from the vaccination is enough to prevent the virus from entering nervous tissue. Thus, the fatal neurological consequences of the disease are averted and the individual only has to recover from the infected bite. This approach is also being used for the treatment of Ebola, one of the fastest and most deadly viruses affecting humans, though usually infecting limited populations. Ebola is also a leading cause of death in gorillas. Transmitted by bats and great apes, this virus can cause death in 70–90 percent of the infected within two weeks. Using newly developed vaccines that boost the immune response, there is hope that immune systems of affected individuals will be better able to control the virus, potentially reducing mortality rates.

Another way of treating viral infections is the use of antiviral drugs. These drugs often have limited ability to cure viral disease but have been used to control and reduce symptoms for a wide variety of viral diseases. For most viruses, these drugs inhibit the virus by blocking the actions of one or more of its proteins. It is important that the targeted proteins be encoded for by viral genes and that these molecules are not present in a healthy host cell. In this way, viral growth is inhibited without damaging the host. There are large numbers of antiviral drugs available to treat infections, some specific for a particular virus and others that can affect multiple viruses.

Antivirals have been developed to treat genital herpes (herpes simplex II) and influenza. For genital herpes, drugs such as acyclovir can reduce the number and duration of the episodes of active viral disease during which patients develop viral lesions in their skins cells. As the virus remains latent in

nervous tissue of the body for life, this drug is not a cure but can make the symptoms of the disease more manageable. For influenza, drugs like Tamiflu can reduce the duration of “flu” symptoms by one or two days, but the drug does not prevent symptoms entirely. Other antiviral drugs, such as Ribavirin, have been used to treat a variety of viral infections.

By far the most successful use of antivirals has been in the treatment of the retrovirus HIV, which causes a disease that, if untreated, is usually fatal within 10–12 years after being infected. Anti-HIV drugs have been able to control viral replication to the point that individuals receiving these drugs survive for a significantly longer time than the untreated.

Anti-HIV drugs inhibit viral replication at many different phases of the HIV replicative cycle. Drugs have been developed that inhibit the fusion of the HIV viral envelope with the plasma membrane of the host cell (fusion inhibitors), the conversion of its RNA genome to double-stranded DNA (reverse transcriptase inhibitors), the integration of the viral DNA into the host genome (integrase inhibitors), and the processing of viral proteins (protease inhibitors).

When any of these drugs are used individually, the virus’ high mutation rate allows the virus to rapidly evolve resistance to the drug. The breakthrough in the treatment of HIV was the development of highly active anti-retroviral therapy (HAART), which involves a mixture of different drugs, sometimes called a drug “cocktail.” By attacking the virus at different stages of its replication cycle, it is difficult for the virus to develop resistance to multiple drugs at the same time. Still, even with the use of combination HAART therapy, there is concern that, over time, the virus will evolve resistance to this therapy. Thus, new anti-HIV drugs are constantly being developed with the hope of continuing the battle against this highly fatal virus.

Section Summary

Viruses are acellular entities that can usually only be seen with an electron microscope. Their genomes contain either DNA or RNA, and they replicate using the replication proteins of a host cell. Viruses are diverse, infecting archaea, bacteria, fungi, plants, and animals. Viruses consist of a nucleic-acid core surrounded by a protein capsid with or without an outer lipid envelope.

Viral replication within a living cell always produces changes in the cell, sometimes resulting in cell death and sometimes slowly killing the infected cells. There are six basic stages in the virus replication cycle: attachment, penetration, uncoating, replication, assembly, and release. A viral infection may be productive, resulting in new virions, or nonproductive, meaning the virus remains inside the cell without producing new virions.

Viruses cause a variety of diseases in humans. Many of these diseases can be prevented by the use of viral vaccines, which stimulate protective immunity against the virus without causing major disease. Viral vaccines may also be used in active viral infections, boosting the ability of the immune system to control or destroy the virus. Antiviral drugs that target enzymes and other protein products of viral genes have been developed and used with mixed success. Combinations of anti-HIV drugs have been used to effectively control the virus, extending the lifespan of infected individuals.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4706#h5p-77

Glossary

acellular: lacking cells
apoptosis: the cell death caused by induction of a cell's own internal mechanisms either as a natural step in the development of a multicellular organism or by other environmental factors such as signals from cells of the immune system
attenuation: the weakening of a virus during vaccine development
capsid: the protein coating of the viral core
cytopathic: causing cell damage
glycoprotein: a protein molecule with attached carbohydrate molecules
vaccine: a weakened solution of virus components, viruses, or other agents that produce an immune response
virion: an individual virus particle outside a host cell
viral envelope: a lipid bilayer that envelops some viruses

12.2 Innate Immunity

Learning Objectives

By the end of this section, you will be able to:

- Describe the body’s innate physical and chemical defenses
- Explain the inflammatory response
- Describe the complement system

The vertebrate, including human, 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 12.8). Innate immunity is not caused by an infection or vaccination and depends initially 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 reencounter a pathogen even years later.

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 12.8 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. The 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 move 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 microorganisms such as bacteria, archaea, and fungi 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

When pathogens 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 cell that engulfs foreign particles and pathogens. Mast cells are produced in the same way as white blood cells, but unlike circulating white blood cells, mast cells take up residence in connective tissues and especially mucosal tissues. They are responsible for releasing chemicals in response to physical injury. They also play a role in the allergic response, which will be discussed later in the chapter.

When a pathogen is recognized as foreign, 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 to 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 12.9). Macrophages follow neutrophils and take over the phagocytosis function and are involved in the resolution of an inflamed site, cleaning up cell debris and pathogens.

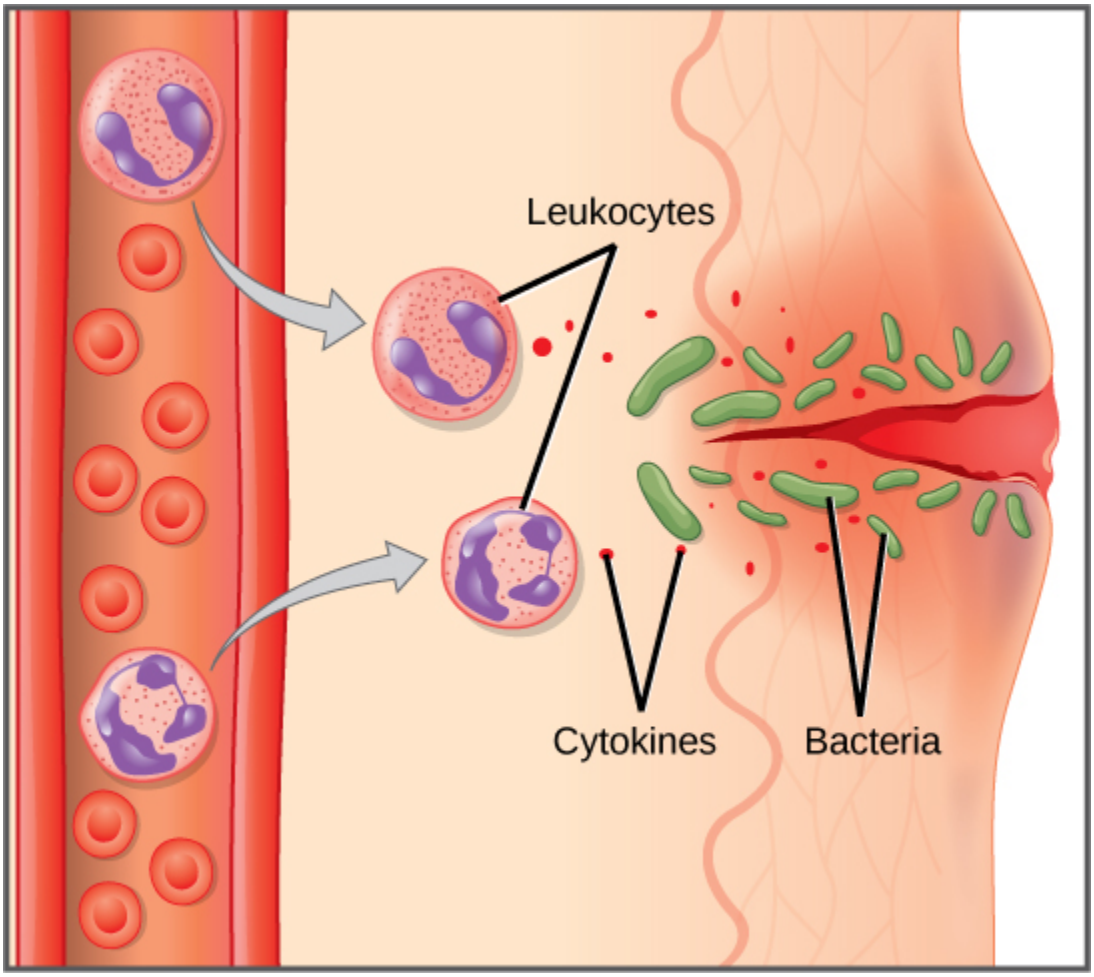


Figure 12.9 White blood cells (leukocytes) release chemicals to stimulate the inflammatory response following a cut in the skin.

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 also 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.

Concept in Action

Check out this 23-second, stop-motion video showing a neutrophil that searches and engulfs fungus spores during an elapsed time of 79 minutes.

Natural Killer Cells

A lymphocyte is a white blood cell that contains a large nucleus (Figure 12.10). Most lymphocytes are associated with the adaptive immune response, but infected cells are identified and destroyed by natural killer cells, the only lymphocytes of the innate immune system. A natural killer (NK) cell is a lymphocyte that can kill cells infected with viruses (or cancerous cells). NK cells identify intracellular infections, especially from viruses, by the altered expression of major histocompatibility class (MHC) I molecules on the surface of infected cells. MHC class I molecules are proteins on the surfaces of all nucleated cells that provide a sample of the cell's internal environment at any given time. Unhealthy cells, whether infected or cancerous, display an altered MHC class I complement on their cell surfaces.

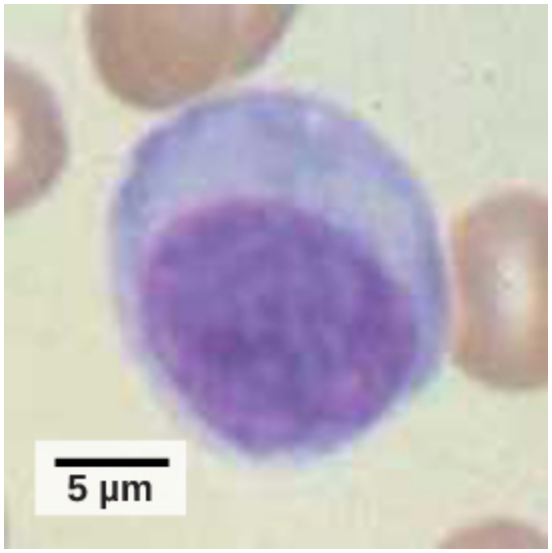


Figure 12.10. Micrograph shows a round cell with a large nucleus occupying more than half of the cell.

After the NK cell detects an infected or tumor cell, it induces programmed cell death, or apoptosis. Phagocytic cells then come along and digest the cell debris left behind. NK cells are constantly patrolling the body and are an effective mechanism for controlling potential infections and preventing cancer progression. The various types of immune cells are shown in Figure 12.11.

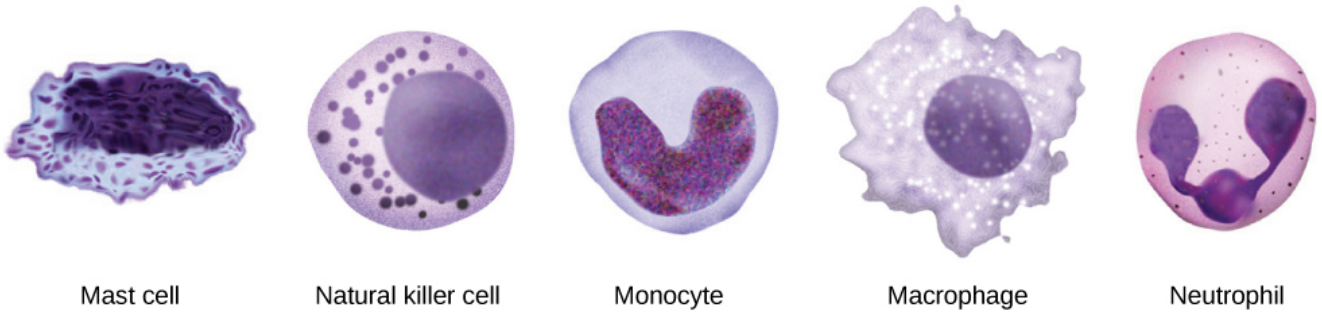


Figure 12.11 Cells involved in the innate immune response include mast cells, natural killer cells, and white blood cells, such as monocytes, macrophages and neutrophils.

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 cell membranes and cause lysis of the cells.

Section Summary

The innate immune system consists first of physical and chemical barriers to infection including the skin and mucous membranes and their secretions, ciliated surfaces, and body hairs. The second line of defense is an internal defense system designed to counter pathogenic threats that bypass the physical and chemical barriers of the body. Using a combination of cellular and molecular responses, the innate immune system identifies the nature of a pathogen and responds with inflammation, phagocytosis, cytokine release, destruction by NK cells, or the complement system.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4713#h5p-78

Glossary

complement system: an array of approximately 20 soluble proteins of the innate immune system that enhance phagocytosis, bore holes in pathogens, and recruit lymphocytes
cytokine: a chemical messenger that regulates cell differentiation, proliferation, and gene expression to effect immune responses
inflammation: the localized redness, swelling, heat, and pain that results from the movement of leukocytes through opened capillaries to a site of infection
innate immunity: an immunity that occurs naturally because of genetic factors or physiology, and is not caused by infection or vaccination
interferon: a cytokine that inhibits viral replication
lymphocyte: a type of white blood cell that includes natural killer cells of the innate immune system and B and T cells of the adaptive immune system
macrophage: a large phagocytic cell that engulfs foreign particles and pathogens
major histocompatibility class (MHC) I: a group of proteins found on the surface of all nucleated cells that signals to immune cells whether the cell is normal or is infected or cancerous; it also provides the appropriate sites into which antigens can be loaded for recognition by lymphocytes
mast cell: a leukocyte that produces inflammatory molecules, such as histamine, in response to large pathogens
monocyte: a type of white blood cell that circulates in the blood and lymph and differentiates into a macrophage after it moves into infected tissue
natural killer (NK) cell: a lymphocyte that can kill cells infected with viruses or tumor cells
neutrophil: a phagocytic leukocyte that engulfs and digests pathogens
white blood cell: a nucleated cell found in the blood that is a part of the immune system; also called leukocytes

12.3 Adaptive Immunity

Learning Objectives

By the end of this section, you will be able to:

- Explain adaptive immunity
- Describe cell-mediated immune response and humoral immune response
- Describe immune tolerance

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 whose surface binding sites are specific to the molecules on the pathogen greatly increase in numbers and attack the invading pathogen. Their attack 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 found in many flat bones, such as the shoulder or pelvic bones. The two types of lymphocytes of the adaptive immune response are B and T cells (Figure 12.12). 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 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 loose in blood and lymph, and T cells are involved in the cell-mediated immune response, which targets infected cells.

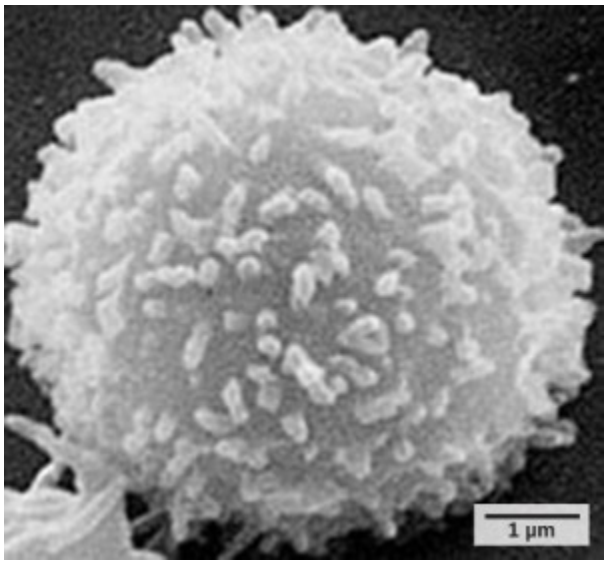


Figure 12.12: This scanning electron micrograph shows a T lymphocyte. T and B cells are indistinguishable by light microscopy but can be differentiated experimentally by probing their surface receptors. (credit: modification of work by NCI; scale-bar data from Matt Russell)

Humoral Immune Response

As mentioned, an antigen is a molecule that stimulates a response in the immune system. Not every molecule is antigenic. B cells participate in a chemical response to antigens present in the body 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. 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 12.13).

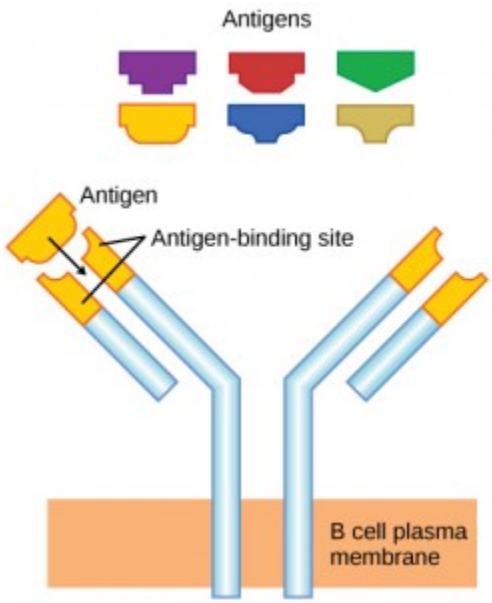


Figure 12.13. B cell receptors are embedded in the membranes of B cells and bind a variety of antigens through their variable regions.

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 and reappears on the surface of the cell bound to an MHC class II molecule. 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 antigen-MHC class II complex and is induced to release cytokines that induce the B cell to divide rapidly, which makes thousands of identical (clonal) cells. These daughter cells become either plasma cells or memory B cells. The memory B cells remain inactive at this point, until another later encounter with the antigen, caused by a reinfection by the same bacteria or virus, results in them dividing into a new population of plasma cells. The plasma cells, on the other hand, produce and secrete large quantities, up to 100 million molecules per hour, of antibody molecules. 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.

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

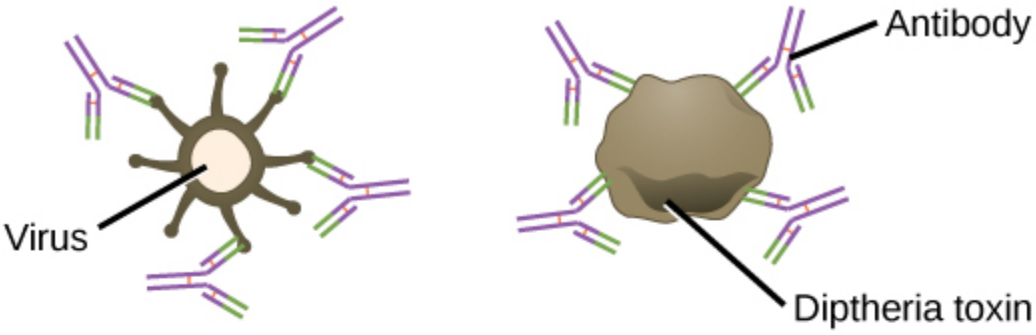
phagocytosis is enhanced when the complexes are present. Finally, antibodies stimulate inflammation, and their presence in mucus and on the skin prevents pathogen attack.

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 12.14). Antibody neutralization can prevent pathogens from entering and infecting host cells. The neutralized antibody-coated pathogens can then be filtered by the spleen and eliminated in urine or feces.

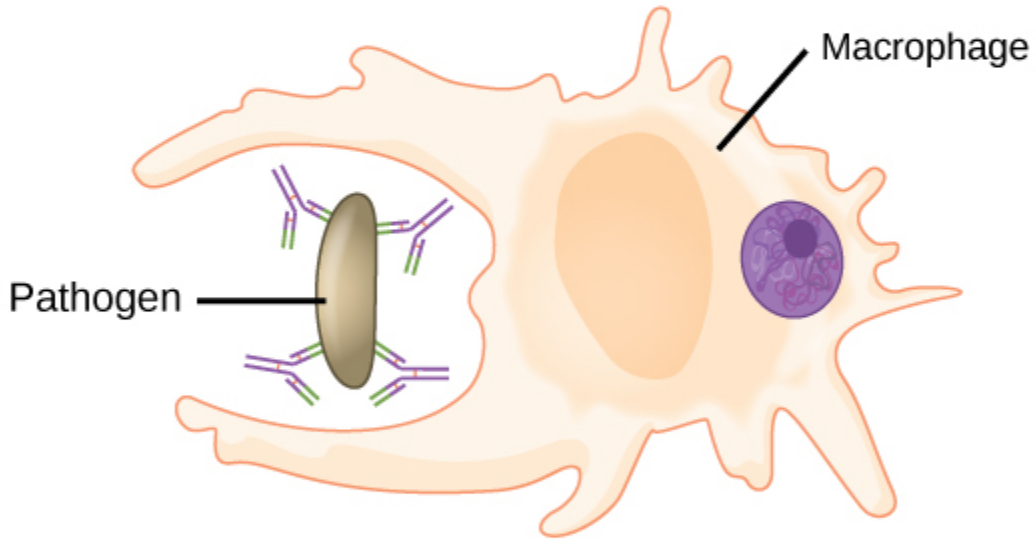
Antibodies also mark pathogens for destruction by phagocytic cells, such as macrophages or neutrophils, in a process called opsonization. In a process called complement fixation, some antibodies provide a place for complement proteins to bind. The combination of antibodies and complement promotes rapid 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 response of the immune system 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.

(a) Neutralization Antibodies prevent a virus or toxic protein from binding their target.



(b) Opsonization A pathogen tagged by antibodies is consumed by a macrophage or neutrophil.



(c) Complement activation Antibodies attached to the surface of a pathogen cell activate the complement system.

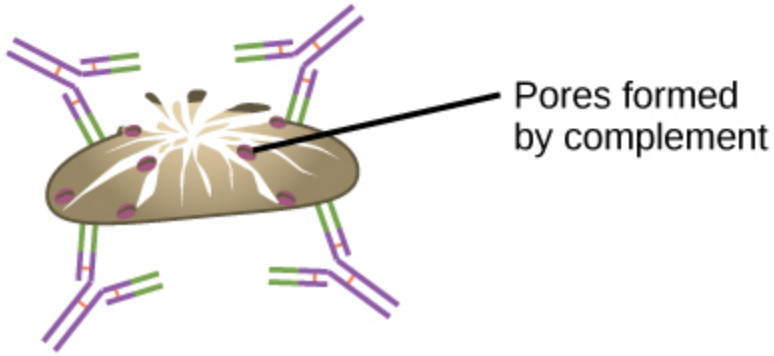


Figure 12.14 Antibodies may inhibit infection by (a) preventing the antigen from binding its

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. After phagocytosis by a macrophage, the phagocytic vesicle fuses with an intracellular lysosome. Within the resulting phagolysosome, the components are broken down into fragments; the fragments are then loaded onto MHC class II molecules and are transported to the cell surface for antigen presentation (Figure 12.15). Helper T cells cannot properly respond to an antigen unless it is processed and embedded in an MHC class II molecule. The APCs express MHC class II on their surfaces, and when combined with a foreign antigen, these complexes signal an invader.

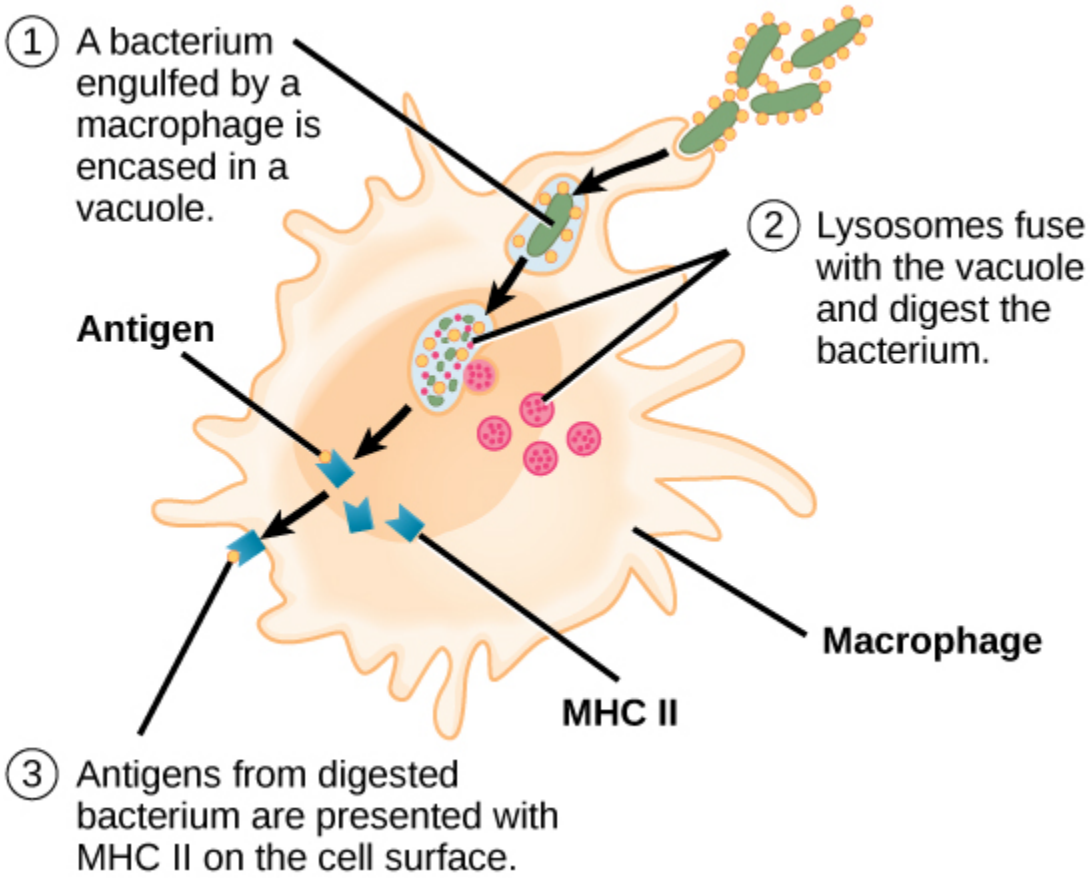


Figure 12.15 An antigen-presenting cell (APC), such as a macrophage, engulfs a foreign antigen, partially digests it in a lysosome, and then embeds it in an MHC class II molecule for presentation at the cell surface. Lymphocytes of the adaptive immune response must interact with antigen-embedded MHC class II molecules to mature into functional immune cells.

Concept in Action

View this animation from Rockefeller University to see how dendritic cells act as sentinels in the body's immune system.

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 12.16.

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.

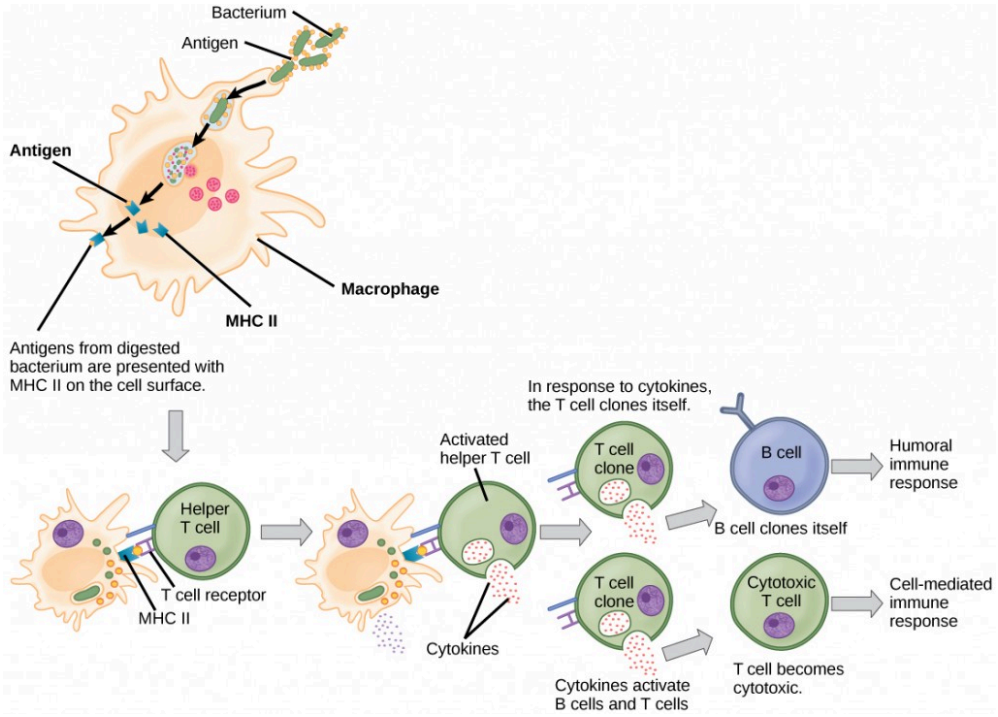


Figure 12.16. A helper T cell becomes activated by binding to an antigen presented by an APC via the MHCII receptor, causing it to release cytokines. Depending on the cytokines released, this activates either the humoral or the cell-mediated immune response.

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 12.17). 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.

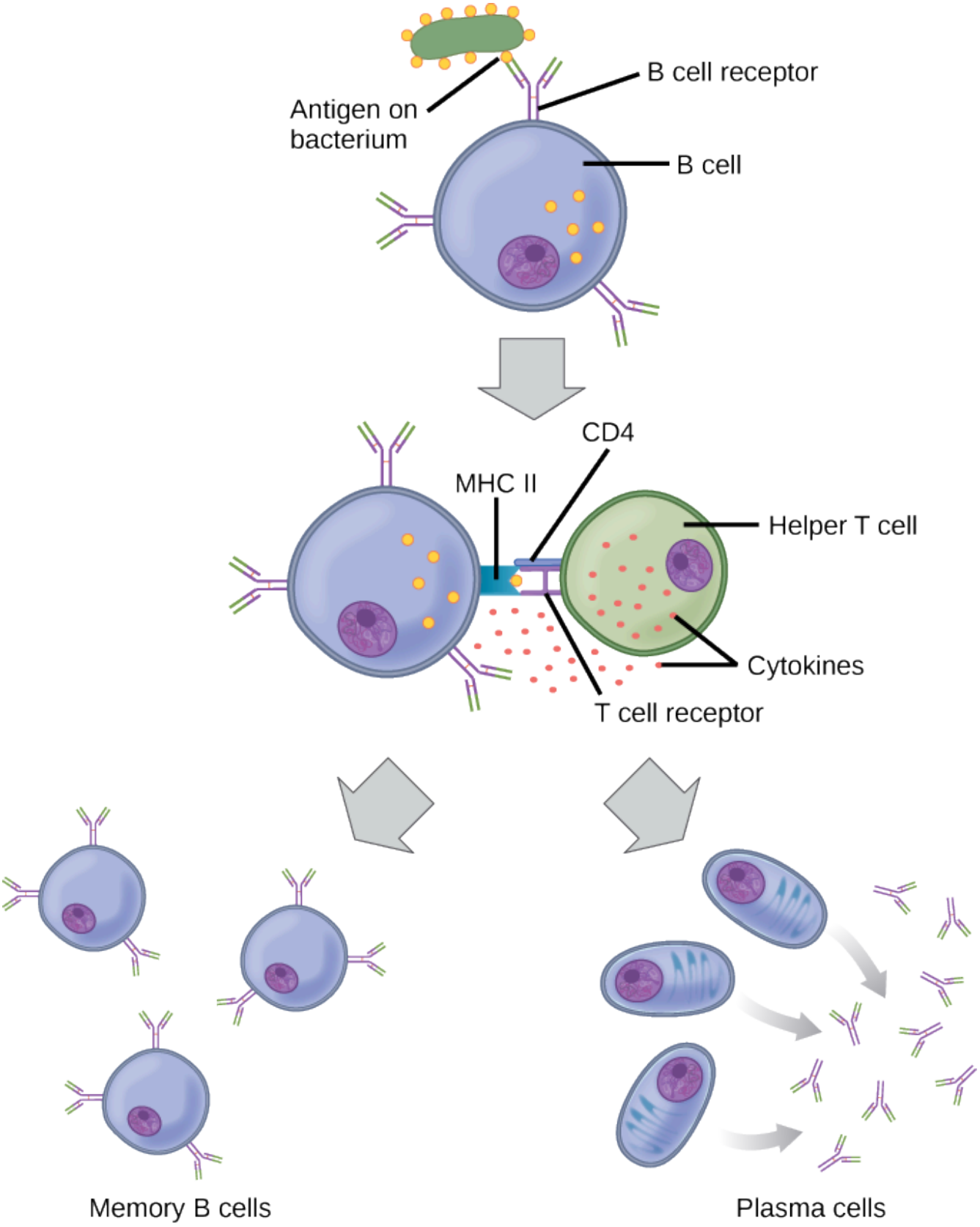


Figure 12.17 After initially binding an antigen to the B cell receptor, a B cell internalizes the antigen and presents it on MHC class II. A helper T cell recognizes the MHC class II- antigen complex and activates the B cell. As a result, memory B cells and plasma cells are made.

The Rh antigen is found on Rh-positive red blood cells. An Rh-negative female can usually carry an Rh-positive fetus to term without difficulty. However, if she has a second Rh-positive fetus, her body may launch an immune attack that causes hemolytic disease of the newborn. Why do you think hemolytic disease is only a problem during the second or subsequent pregnancies?

<!– If the blood of the mother and fetus mixes, memory cells that recognize the Rh antigen of the fetus can form in the mother late in the first pregnancy. During subsequent pregnancies, these memory cells launch an immune attack on the fetal blood cells of an Rh-positive fetus. Injection of anti-Rh antibody during the first pregnancy prevents the immune response from occurring.–>

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 output tens to hundreds-fold greater antibody amounts than were secreted during the primary response (Figure 12.18). 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.

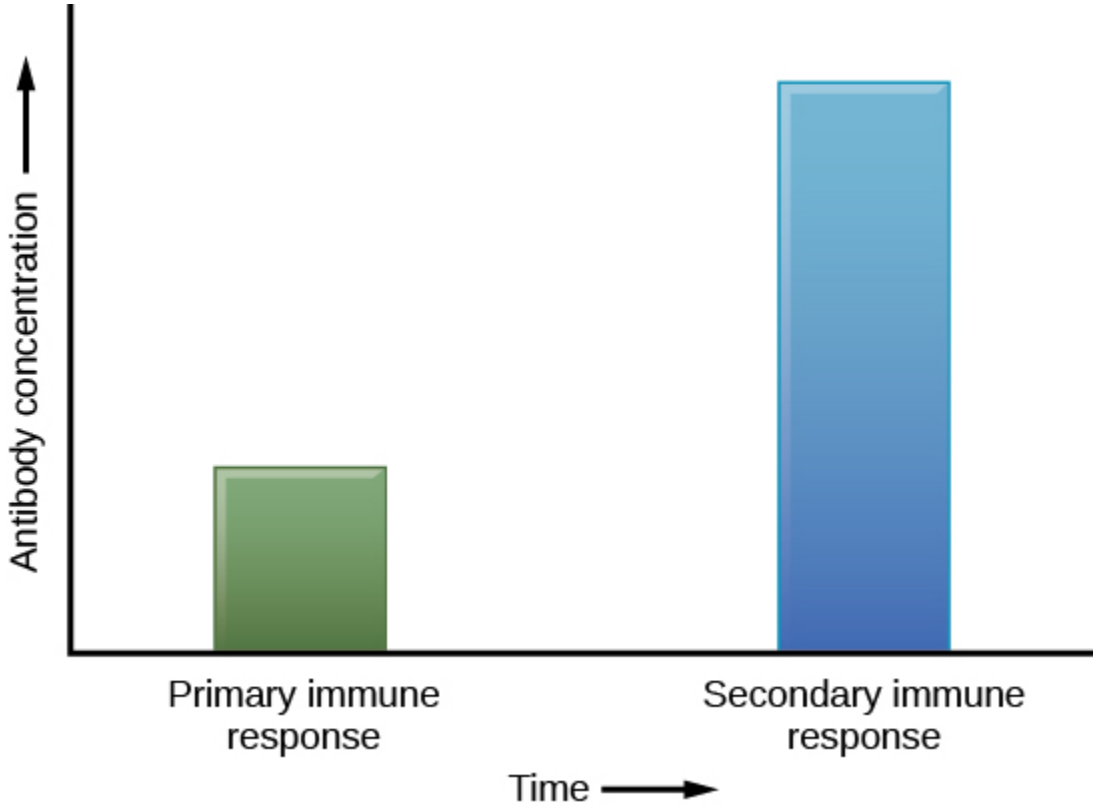


Figure 12.18 In the primary response to infection, antibodies are secreted first from plasma cells. Upon re-exposure to the same pathogen, memory cells differentiate into antibody-secreting plasma cells that output a greater amount of antibody for a longer period of time.

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 12.19). 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.

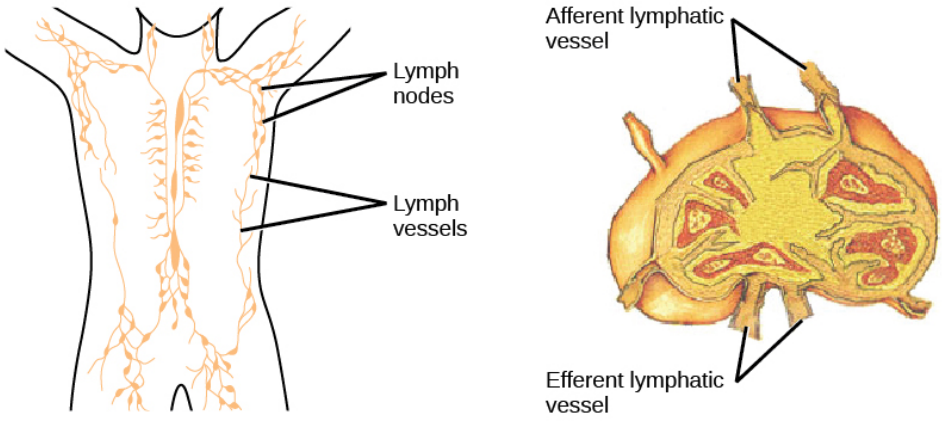


Figure 12.19. (a) Lymphatic vessels carry a clear fluid called lymph throughout the body. The liquid passes through (b) lymph nodes that filter the lymph that enters the node through afferent vessels and leaves through efferent vessels; lymph nodes are filled with lymphocytes that purge infecting cells. (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 12.20). 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.

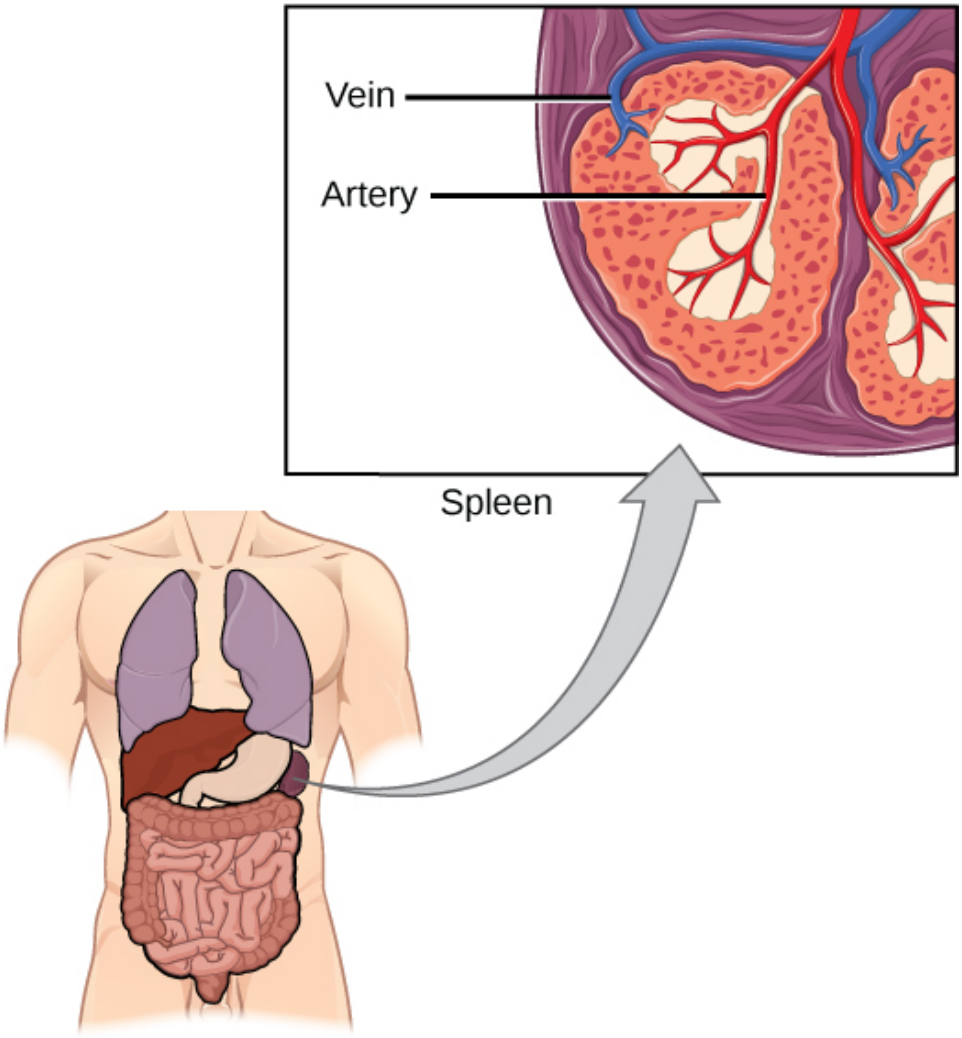


Figure 12.20 The spleen functions to immunologically filter the blood and allow for communication between cells corresponding to the innate and adaptive immune responses. (credit: modification of work by NCI, NIH)

Mucosal Immune System

The innate and adaptive immune responses compose the systemic immune system (affecting the whole body), which is distinct from the mucosal immune system. Mucosa associated lymphoid tissue (MALT) is a crucial component of a functional immune system because mucosal surfaces, such as the nasal passages, are the first tissues onto which inhaled or ingested pathogens are deposited. The mucosal tissue includes the mouth, pharynx, and esophagus, and the gastrointestinal, respiratory, and urogenital tracts.

Mucosal immunity is formed by MALT, which functions independently of the systemic immune system, and which has its own innate and adaptive components. MALT is a collection of lymphatic tissue that combines with epithelial tissue lining the mucosa throughout the body. This tissue functions as the immune barrier and response in areas of the body with direct contact to the external environment. The systemic and mucosal immune systems use many of the same cell types. Foreign particles that

make their way to MALT are taken up by absorptive epithelial cells and delivered to APCs located directly below the mucosal tissue. APCs of the mucosal immune system are primarily dendritic cells, with B cells and macrophages having minor roles. Processed antigens displayed on APCs are detected by T cells in the MALT and at the tonsils, adenoids, appendix, or the mesenteric lymph nodes of the intestine. Activated T cells then migrate through the lymphatic system and into the circulatory system to mucosal sites of infection.

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.

Section Summary

The adaptive immune response is a slower-acting, longer-lasting, and more specific response than the innate response. However, the adaptive response requires information from the innate immune system to function. APCs display antigens on MHC molecules to naïve T cells. T cells with cell-surface receptors that bind a specific antigen will bind to that APC. In response, the T cells differentiate and proliferate, becoming T_H cells or T_C cells. T_H cells stimulate B cells that have engulfed and presented pathogen-derived antigens. B cells differentiate into plasma cells that secrete antibodies, whereas T_C cells destroy infected or cancerous cells. Memory cells are produced by activated and proliferating B and T cells and persist after a primary exposure to a pathogen. If re-exposure occurs, memory cells differentiate into effector cells without input from the innate immune system. The mucosal immune system is largely independent of the systemic immune system but functions in parallel to protect the extensive mucosal surfaces of the body. Immune tolerance is brought about by T_{reg} cells to limit reactions to harmless antigens and the body's own molecules.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4725#h5p-79

Glossary

active immunity: an immunity that occurs as a result of the activity of the body's own cells rather than from antibodies acquired from an external source

adaptive immunity: a specific immune response that occurs after exposure to an antigen either from a pathogen or a vaccination

antibody: a protein that is produced by plasma cells after stimulation by an antigen; also known as an immunoglobulin

antigen: a macromolecule that reacts with cells of the immune system and which may or may not have a stimulatory effect

antigen-presenting cell (APC): an immune cell that detects, engulfs, and informs the adaptive immune response about an infection by presenting the processed antigen on its cell surface

B cell: a lymphocyte that matures in the bone marrow

cell-mediated immune response: an adaptive immune response that is controlled by T cells

cytotoxic T lymphocyte (T_C): an adaptive immune cell that directly kills infected cells via enzymes, and that releases cytokines to enhance the immune response

dendritic cell: an immune cell that processes antigen material and presents it on the surface of its cell in MHC class II molecules and induces an immune response in other cells

effector cell: a lymphocyte that has differentiated, such as a B cell, plasma cell, or cytotoxic T cell

helper T lymphocyte (T_H): a cell of the adaptive immune system that binds APCs via MHC class II molecules and stimulates B cells or secretes cytokines to initiate the immune response

humoral immune response: the adaptive immune response that is controlled by activated B cells and antibodies

immune tolerance: an acquired ability to prevent an unnecessary or harmful immune response to a detected foreign body known not to cause disease

lymph: the watery fluid present in the lymphatic circulatory system that bathes tissues and organs with protective white blood cells and does not contain erythrocytes

memory cell: 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

major histocompatibility class (MHC) II molecule: a protein found on the surface of antigen-presenting cells that signals to immune cells whether the cell is normal or is infected or cancerous; it provides the appropriate template into which antigens can be loaded for recognition by lymphocytes

passive immunity: an immunity that does not result from the activity of the body's own immune cells but by transfer of antibodies from one individual to another

primary immune response: the response of the adaptive immune system to the first exposure to an antigen

secondary immune response: the response of the adaptive immune system to a second or later exposure to an antigen mediated by memory cells

T cell: a lymphocyte that matures in the thymus gland

12.4 Disruptions in the Immune System

Learning Objectives

By the end of this section, you will be able to:

- Describe hypersensitivity
- Define autoimmunity

A functioning immune system is essential for survival, but even the sophisticated cellular and molecular defenses of the mammalian immune response can be defeated by pathogens at virtually every step. In the competition between immune protection and pathogen evasion, pathogens have the advantage of more rapid evolution because of their shorter generation time, large population sizes and often higher mutation rates. Thus pathogens have evolved a diverse array of immune escape mechanisms. For instance, *Streptococcus pneumoniae* (the bacterium that causes pneumonia and meningitis) surrounds itself with a capsule that inhibits phagocytes from engulfing it and displaying antigens to the adaptive immune system. *Staphylococcus aureus* (the bacterium that can cause skin infections, abscesses, and meningitis) synthesizes a toxin called leukocidin that kills phagocytes after they engulf the bacterium. Other pathogens can also hinder the adaptive immune system. HIV infects T_H cells using their CD4 surface molecules, gradually depleting the number of T_H cells in the body (Figure 12.21); this inhibits the adaptive immune system's capacity to generate sufficient responses to infection or tumors. As a result, HIV-infected individuals often suffer from infections that would not cause illness in people with healthy immune systems but which can cause devastating illness to immune-compromised individuals.

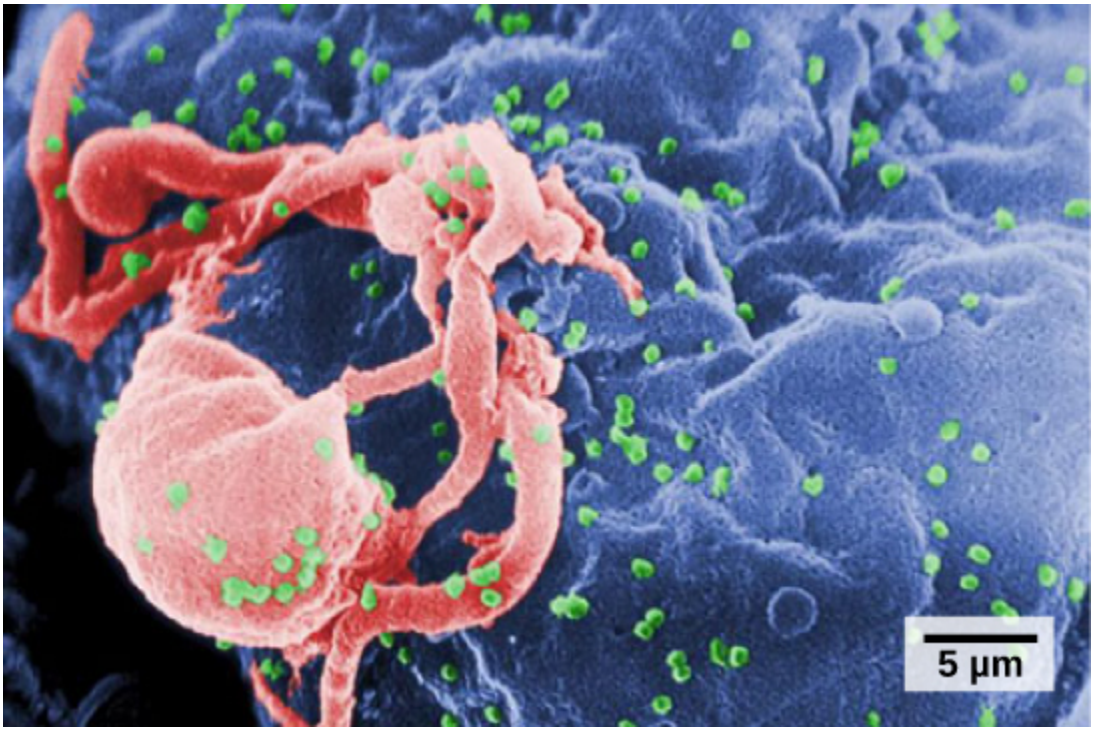


Figure 12.21 HIV (green) is shown budding from a lymphocyte cell (red) in culture. (credit: modification of work by C. Goldsmith, CDC; scale-bar data from Matt Russell)

Inappropriate responses of immune cells and molecules themselves can also disrupt the proper functioning of the entire system, leading to host-cell damage that can become fatal.

Immunodeficiency

Immunodeficiency is a failure, insufficiency, or delay in the response of the immune system, which may be acquired or inherited. Immunodeficiency can allow pathogens or tumor cells to gain a foothold and replicate or proliferate to high enough levels so that the immune system becomes overwhelmed. Immunodeficiency can be acquired as a result of infection with certain pathogens that attack the cells of the immune system itself (such as HIV), chemical exposure (including certain medical treatments such as chemotherapy), malnutrition, or extreme stress. For instance, radiation exposure can destroy populations of lymphocytes and elevate an individual's susceptibility to infections and cancer. Rarely, primary immunodeficiencies that are present from birth may also occur. For example, severe combined immunodeficiency disease (SCID) is a condition in which children are born without functioning B or T cells.

Hypersensitivities

A maladaptive immune response toward harmless foreign substances or self-antigens that occur after tissue sensitization is termed a hypersensitivity. Types of hypersensitivities include immediate, delayed,

and autoimmune. A large proportion of the human population is affected by one or more types of hypersensitivity.

Allergies

The immune reaction that results from immediate hypersensitivities in which an antibody-mediated immune response occurs within minutes of exposure to a usually harmless antigen is called an allergy. In the United States, 20 percent of the population exhibits symptoms of allergy or asthma, whereas 55 percent test positive against one or more allergens. On initial exposure to a potential allergen, an allergic individual synthesizes antibodies through the typical process of APCs presenting processed antigen to T_H cells that stimulate B cells to produce the antibodies. The antibody molecules interact with mast cells embedded in connective tissues. This process primes, or sensitizes, the tissue. On subsequent exposure to the same allergen, antibody molecules on mast cells bind the antigen and stimulate the mast cell to release histamine and other inflammatory chemicals; these chemical mediators then recruit eosinophils (a type of white blood cell), which also appear to be adapted to responding to parasitic worms (Figure 12.22). Eosinophils release factors that enhance the inflammatory response and the secretions of mast cells. The effects of an allergic reaction range from mild symptoms like sneezing and itchy, watery eyes to more severe or even life-threatening reactions involving intensely itchy welts or hives, airway constriction with severe respiratory distress, and plummeting blood pressure caused by dilating blood vessels and fluid loss from the circulatory system. This extreme reaction, typically in response to an allergen introduced to the circulatory system, is known as anaphylactic shock. Antihistamines are an insufficient counter to anaphylactic shock and if not treated with epinephrine to counter the blood pressure and breathing effects, this condition can be fatal.

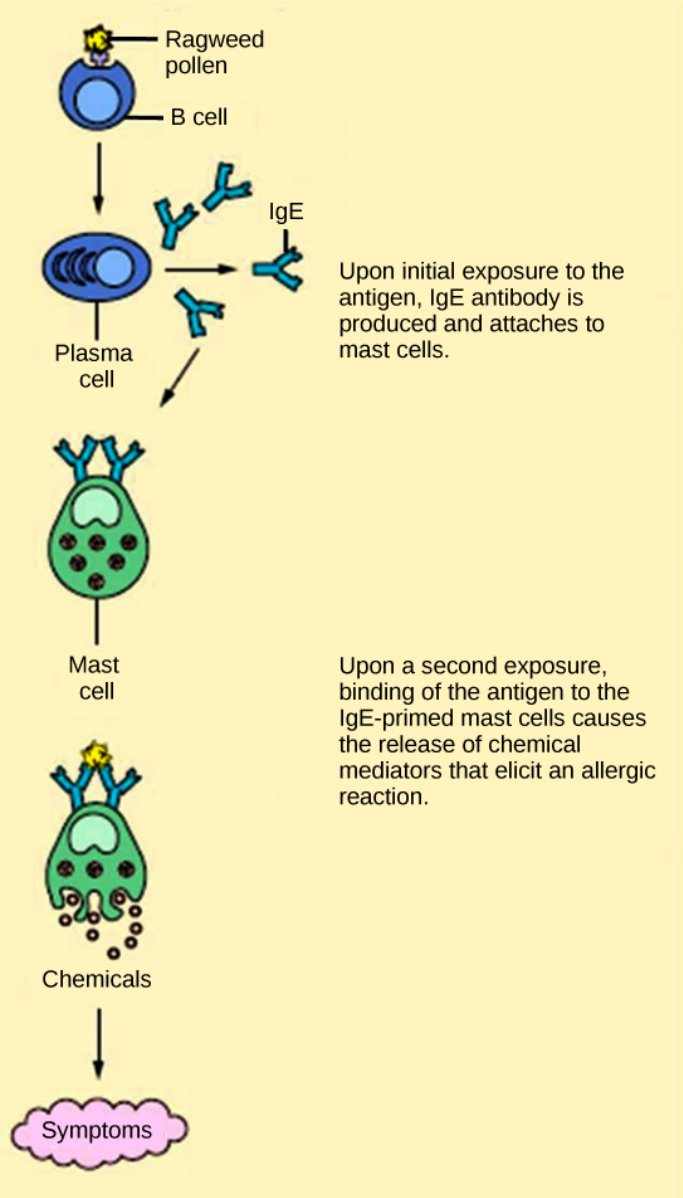


Figure 12.22 On first exposure to an allergen, an antibody is synthesized by plasma cells in response to a harmless antigen. The antibodies bind to mast cells, and on secondary exposure, the mast cells release histamines and other modulators that cause the symptoms of allergy. (credit: modification of work by NIH)

Delayed hypersensitivity is a cell-mediated immune response that takes approximately one to two days after secondary exposure for a maximal reaction. This type of hypersensitivity involves the T_H1 cytokine-mediated inflammatory response and may cause local tissue lesions or contact dermatitis (rash or skin irritation). Delayed hypersensitivity occurs in some individuals in response to contact with certain types of jewelry or cosmetics. Delayed hypersensitivity facilitates the immune response to poison ivy and is also the reason why the skin test for tuberculosis results in a small region of

inflammation on individuals who were previously exposed to *Mycobacterium tuberculosis*, the organism that causes tuberculosis.

Concept in Action

Try your hand at diagnosing an allergic reaction by selecting one of the interactive case studies at the World Allergy Organization website.

Autoimmunity

Autoimmunity is a type of hypersensitivity to self-antigens that affects approximately five percent of the population. Most types of autoimmunity involve the humoral immune response. An antibody that inappropriately marks self-components as foreign is termed an autoantibody. In patients with myasthenia gravis, an autoimmune disease, muscle-cell receptors that induce contraction in response to acetylcholine are targeted by antibodies. The result is muscle weakness that may include marked difficultly with fine or gross motor functions. In systemic lupus erythematosus, a diffuse autoantibody response to the individual's own DNA and proteins results in various systemic diseases (Figure 12.23). Systemic lupus erythematosus may affect the heart, joints, lungs, skin, kidneys, central nervous system, or other tissues, causing tissue damage through antibody binding, complement recruitment, lysis, and inflammation.

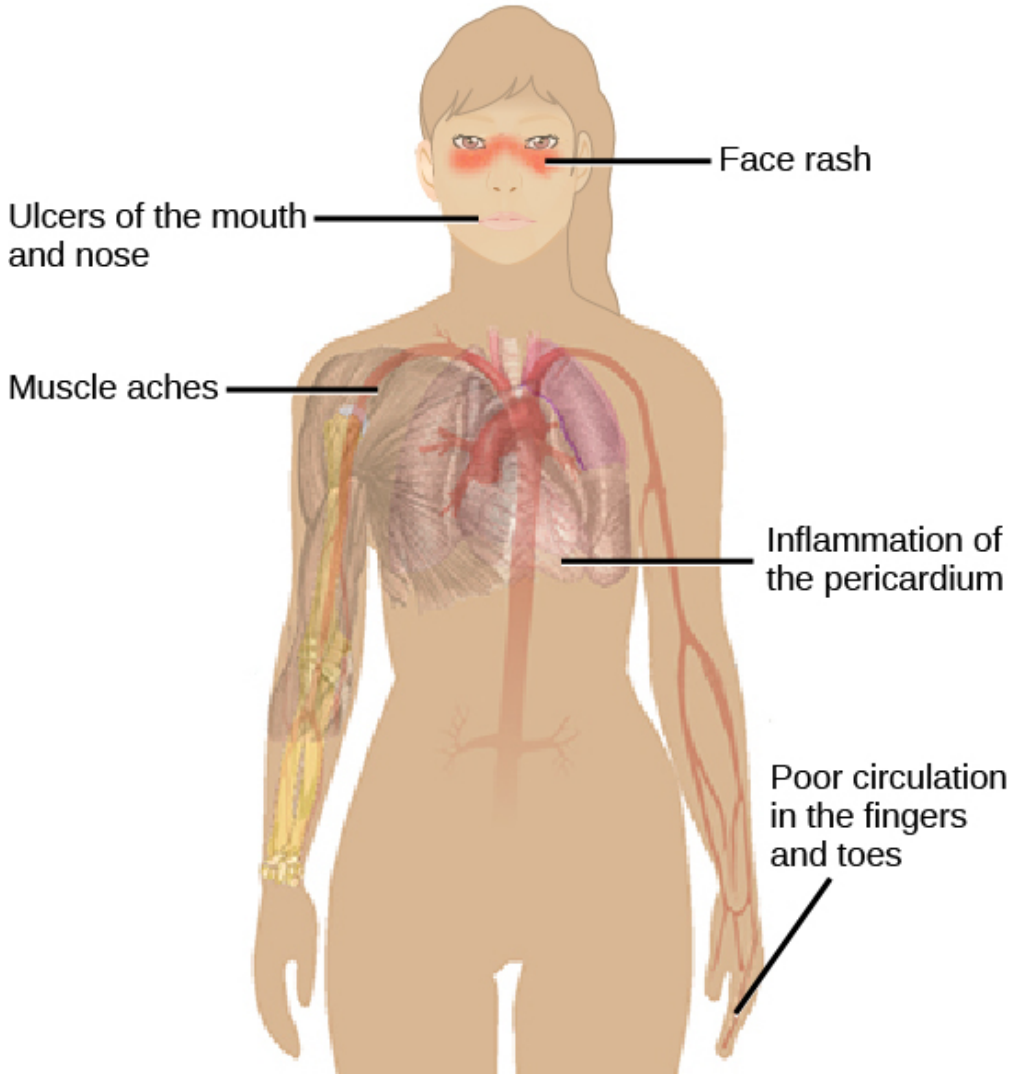


Figure 12.23 Systemic lupus erythematosus is characterized by autoimmunity to the individual's own DNA and/or proteins, which leads to varied dysfunction of the organs. (credit: modification of work by Mikael Häggström)

Autoimmunity can develop with time and its causes may be rooted in molecular mimicry, a situation in which one molecule is similar enough in shape to another molecule that it binds the same immune receptors. Antibodies and T-cell receptors may bind self-antigens that are structurally similar to pathogen antigens. As an example, infection with *Streptococcus pyogenes* (the bacterium that causes strep throat) may generate antibodies or T cells that react with heart muscle, which has a similar structure to the surface of *S. pyogenes*. These antibodies can damage heart muscle with autoimmune attacks, leading to rheumatic fever. Insulin-dependent (Type 1) diabetes mellitus arises from a destructive inflammatory T_H1 response against insulin-producing cells of the pancreas. Patients with this autoimmunity must be treated with regular insulin injections.

Section Summary

Immune disruptions may involve insufficient immune responses or inappropriate immune responses. Immunodeficiency increases an individual's susceptibility to infections and cancers. Hypersensitivities are misdirected responses either to harmless foreign particles, as in the case of allergies, or to the individual's own tissues, as in the case of autoimmunity. Reactions to self-components may be the result of molecular mimicry.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://opentextbc.ca/biology/?p=4731#h5p-80

Glossary

allergy: an immune reaction that results from immediate hypersensitivities in which an antibody-mediated immune response occurs within minutes of exposure to a harmless antigen
autoantibody: an antibody that incorrectly marks "self" components as foreign and stimulates the immune response
autoimmunity: a type of hypersensitivity to self-antigens
hypersensitivity: a spectrum of inappropriate immune responses toward harmless foreign particles or self-antigens; occurs after tissue sensitization and includes immediate-type (allergy), delayed-type, and autoimmunity
immunodeficiency: a failure, insufficiency, or delay at any level of the immune system, which may be acquired or inherited

Chapter 13: Introduction to Animal Reproduction and Development

Figure 13.1 Female seahorses produce eggs that are then fertilized by the male. Unlike with almost all other animals, the young then develop in a pouch of the male seahorse until birth. (credit: "cliff1066"/Flickr)

In the animal kingdom, each species has its unique adaptations for reproduction. Asexual reproduction produces genetically identical offspring (clones), whereas in sexual reproduction, the genetic material of two individuals combines to produce offspring that are genetically different from their parents. During sexual reproduction the male gamete (sperm) may be placed inside the female's body for internal fertilization, the sperm may be left in the environment for the female to pick up and place in her body, or both sperm and eggs may be released into the environment for external fertilization. Seahorses provide an example of the latter, but with a twist (Figure 13.1). Following a mating dance, the female releases eggs into the male seahorse's abdominal brood pouch and the male releases sperm into the water, which then find their way into the brood pouch to fertilize the eggs. The fertilized eggs develop in the pouch for several weeks.

13.1 How Animals Reproduce

Learning Objectives

By the end of this section, you will be able to:

- Describe advantages and disadvantages of asexual and sexual reproduction
- Discuss asexual reproduction methods
- Discuss sexual reproduction methods
- Discuss internal and external methods of fertilization

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. 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, species of turbellarian flatworms commonly called the planarians, 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 13.2), 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 13.2 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 13.3).

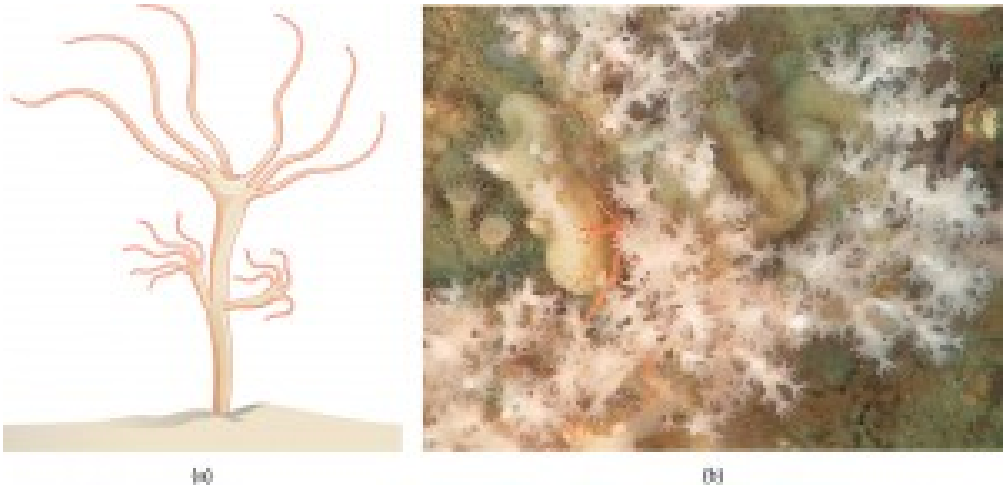


Figure 13.3 (a) Hydra reproduce asexually through budding: a bud forms on the tubular body of an adult hydra, develops a mouth and tentacles, and then detaches from its parent. The new hydra is fully developed and will find its own location for attachment. (b) Some coral, such as the Lophelia pertusa shown here, can reproduce through budding. (credit b: modification of work by Ed Bowlby, NOAA/Olympic Coast NMS; NOAA/OAR/Office of Ocean Exploration) Part a: This shows a hydra, which has a stalk-like body with tentacles growing out the top. A smaller hydra is budding from the side of the stalk. Part b: This photo shows branching white coral polyps.

Concept in Action

View this video to see a hydra budding.

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 13.4) 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.

Figure 13.4 (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: modification 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 in the 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—such as certain 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 when the females have been isolated from males. It is possible that the asexual reproduction observed occurred in response to unusual circumstances and would normally not occur.

Sexual Reproduction

Sexual reproduction is the combination of reproductive cells from two individuals to form 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. Sex determination, the mechanism that determines which sex an individual develops into, also can vary.

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 13.5) 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.

Figure 13.5 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)

Sex Determination

Mammalian sex is determined genetically by the combination of X and Y chromosomes. Individuals homozygous for X (XX) are female and heterozygous individuals (XY) are male. In mammals, the presence of a Y chromosome causes the development of male characteristics and its absence results in female characteristics. The XY system is also found in some insects and plants.

Bird sex determination is dependent on the combination of Z and W chromosomes. Homozygous for Z (ZZ) results in a male and heterozygous (ZW) results in a female. Notice that this system is the opposite of the mammalian system because in birds the female is the sex with the different sex chromosomes. The W appears to be essential in determining the sex of the individual, similar to the Y chromosome in mammals. Some fish, crustaceans, insects (such as butterflies and moths), and reptiles use the ZW system.

More complicated chromosomal sex determining systems also exist. For example, some swordtail fish have three sex chromosomes in a population.

The sex of some other species is not determined by chromosomes, but by some aspect of the environment. Sex determination in alligators, some turtles, and tuataras, for example, is dependent on the temperature during the middle third of egg development. This is referred to as environmental sex determination, or more specifically, as temperature-dependent sex determination. In many turtles, cooler temperatures during egg incubation produce males and warm temperatures produce females, while in many other species of turtles, the reverse is true. In some crocodiles and some turtles, moderate temperatures produce males and both warm and cool temperatures produce females.

Individuals of some species change their sex during their lives, switching from one to the other. If the individual is female first, it is termed protogyny or “first female,” if it is male first, it is termed protandry or “first male.” Oysters are born male, grow in size, and become female and lay eggs. The wrasses, a family of reef fishes, are all sequential hermaphrodites. Some of these species live in closely coordinated schools with a dominant male and a large number of smaller females. If the male dies, a female increases in size, changes sex, and becomes the new dominant male.

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 13.6).

Figure 13.6 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 13.7 **a**). This occurs in 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 13.7 **b**), 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 13.7 **c**), some cartilaginous fish, and a few reptiles.

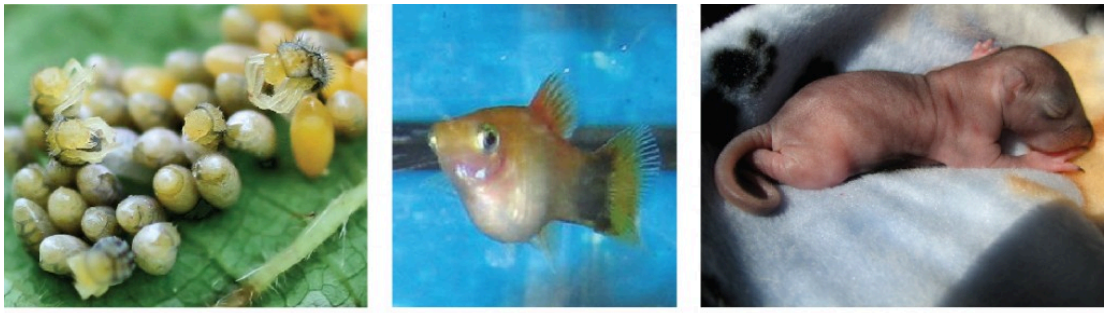


Figure 13.7 In (a) oviparity, young develop in eggs outside the female body, as with these Harmonia axydridis beetles hatching. Some aquatic animals, like this (b) pregnant Xiphophorus maculatus are ovoviparous, with the egg developing inside the female and nutrition supplied primarily from the yolk. 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)

Section Summary

Reproduction may be asexual when one individual produces genetically identical offspring, or sexual when the genetic material from two individuals is combined to produce genetically diverse offspring. Asexual reproduction in animals occurs through fission, budding, fragmentation, and parthenogenesis.

Sexual reproduction may involve fertilization inside the body or in the external environment. A species may have separate sexes or combined sexes; when the sexes are combined they may be expressed at different times in the life cycle. The sex of an individual may be determined by various chromosomal systems or environmental factors such as temperature.

Sexual reproduction starts with the combination of a sperm and an egg in a process called fertilization. This can occur either outside the bodies or inside the female. The method of fertilization varies among animals. Some species release the egg and sperm into the environment, some species retain the egg and receive the sperm into the female body and then expel the developing embryo covered with shell, while still other species retain the developing offspring throughout the gestation period.

Exercises

1. In which group is parthenogenesis a normal event?
    1. chickens
    2. bees
    3. rabbits
    4. sea stars
2. Genetically unique individuals are produced through _____.
    1. sexual reproduction
    2. parthenogenesis
    3. budding
    4. fragmentation
3. External fertilization occurs in which type of environment?
    1. aquatic
    2. forested
    3. savanna
    4. steppe
4. What might be a disadvantage to temperature-dependent sex determination?
5. Compared to separate sexes and assuming self-fertilizing is not possible, what might be one advantage and one disadvantage to hermaphroditism?

Answers

1. B
2. A
3. A
4. Temperatures can vary from year to year and an unusually cold or hot year might produce offspring all of one sex, making it hard for individuals to find mates.

5. A possible advantage of hermaphroditism might be that anytime an individual of the same species is encountered a mating is possible, unlike separate sexes that must find an individual of the right sex to mate. (Also, every individual in a hermaphrodite population is able to produce offspring, which is not the case in populations with separate sexes.) A disadvantage might be that hermaphrodite populations are less efficient because they do not specialize in one sex or another, which means a hermaphrodite does not produce as many offspring through eggs or sperm as do species with separate sexes. (Other answers are possible.)

Glossary

asexual reproduction: a mechanism that produces offspring that are genetically identical to the parent
budding: a form of asexual reproduction that results from the outgrowth of a part of an organism leading to a separation from the original animal into two individuals
external fertilization: the fertilization of eggs by sperm outside an animal's body, often during spawning
fission: (also, binary fission) a form of asexual reproduction in which an organism splits into two separate organisms or two parts that regenerate the missing portions of the body
fragmentation: the breaking of an organism into parts and the growth of a separate individual from each part
hermaphroditism: the state of having both male and female reproductive structures within the same individual
internal fertilization: the fertilization of eggs by sperm inside the body of the female
oviparity: a process by which fertilized eggs are laid outside the female's body and develop there, receiving nourishment from the yolk that is a part of the egg
ovoviparity: a process by which fertilized eggs are retained within the female; the embryo obtains its nourishment from the egg's yolk, and the young are fully developed when they are hatched
parthenogenesis: a form of asexual reproduction in which an egg develops into a complete individual without being fertilized
sex determination: the mechanism by which the sex of individuals in sexually reproducing organisms is initially established
sexual reproduction: a form of reproduction in which cells containing genetic material from two individuals combines to produce genetically unique offspring
viviparity: a process in which the young develop within the female and are born in a nonembryonic state

13.2 Development and Organogenesis

Learning Objectives

By the end of this section, you will be able to:

- Explain how the embryo forms from the zygote
- Discuss the role of cleavage and gastrulation in animal development
- Describe organogenesis

The process by which an organism develops from a single-celled zygote to a multi-cellular organism is complex and well regulated. The regulation occurs through signaling between cells and tissues and responses in the form of differential gene expression.

Early Embryonic Development

Fertilization is the process in which gametes (an egg and sperm) fuse to form a zygote (Figure 13.8). To ensure that the offspring has only one complete diploid set of chromosomes, only one sperm must fuse with one egg. In mammals, a layer called the zona pellucida protects the egg. At the tip of the head of a sperm cell is a structure like a lysosome called the acrosome, which contains enzymes. When a sperm binds to the zona pellucida, a series of events, called the acrosomal reactions, take place. These reactions, involving enzymes from the acrosome, allow the sperm plasma membrane to fuse with the egg plasma membrane and permit the sperm nucleus to transfer into the ovum. The nuclear membranes of the egg and sperm break down and the two haploid nuclei fuse to form a diploid nucleus or genome.

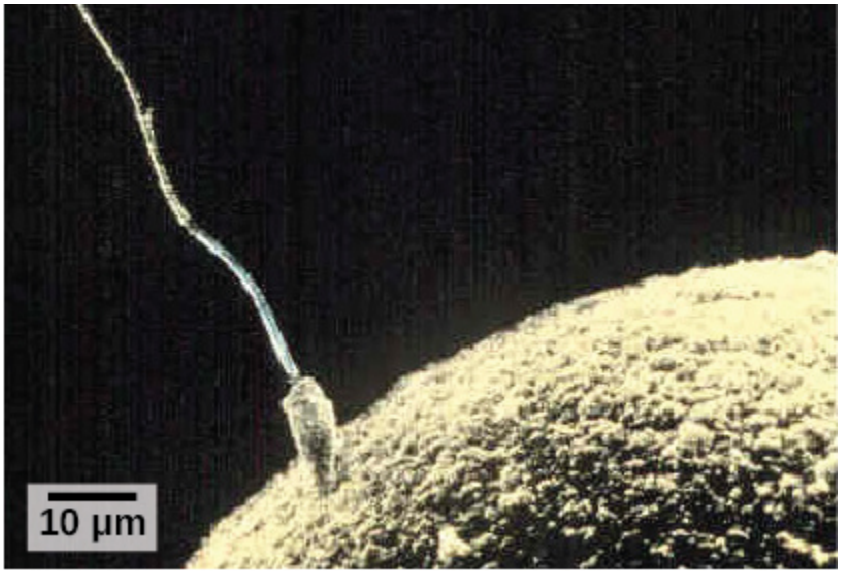


Figure 13.8 Fertilization is the process in which sperm and egg fuse to form a zygote. (credit: scale-bar data from Matt Russell)

To ensure that no more than one sperm fertilizes the egg, once the acrosomal reactions take place at one location of the egg membrane, the egg releases proteins in other locations to prevent other sperm from fusing with the egg.

The development of multi-cellular organisms begins from this single-celled zygote, which undergoes rapid cell division, called cleavage (Figure 13.9 **a**), to form a hollow ball of cells called a blastula (Figure 13.9 **b**).

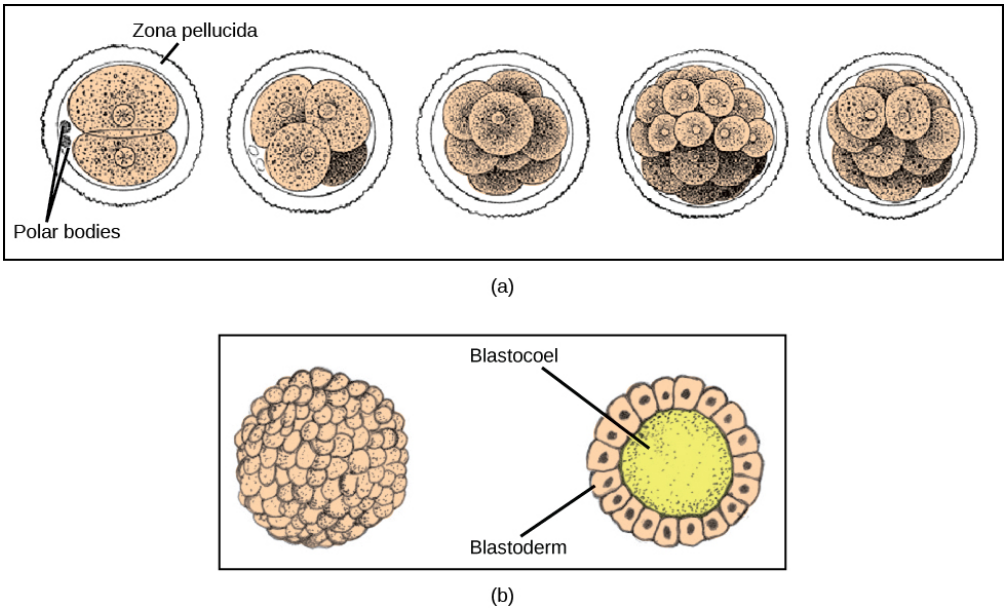


Figure 13.9 (a) During cleavage, the zygote rapidly divides into multiple cells. (b) The cells rearrange themselves to form a hollow ball called the blastula. (credit a: modification of work by Gray's Anatomy; credit b: modification of work by Pearson Scott Foresman; donated to the Wikimedia Foundation)

In mammals, the blastula forms the blastocyst in the next stage of development. Here the cells in the

blastula arrange themselves in two layers: the inner cell mass, and an outer layer called the trophoblast. The inner cell mass will go on to form the embryo. The trophoblast secretes enzymes that allow implantation of the blastocyst into the endometrium of the uterus. The trophoblast will contribute to the placenta and nourish the embryo.

Concept in Action

Visit the Virtual Human Embryo project at the Endowment for Human Development site to click through an interactive of the stages of embryo development, including micrographs and rotating 3-D images.

The cells in the blastula then rearrange themselves spatially to form three layers of cells. This process is called gastrulation. During gastrulation, the blastula folds in on itself and cells migrate to form the three layers of cells (Figure 13.10) in a structure, the gastrula, with a hollow space that will become the digestive tract. Each of the layers of cells is called a germ layer and will differentiate into different organ systems.

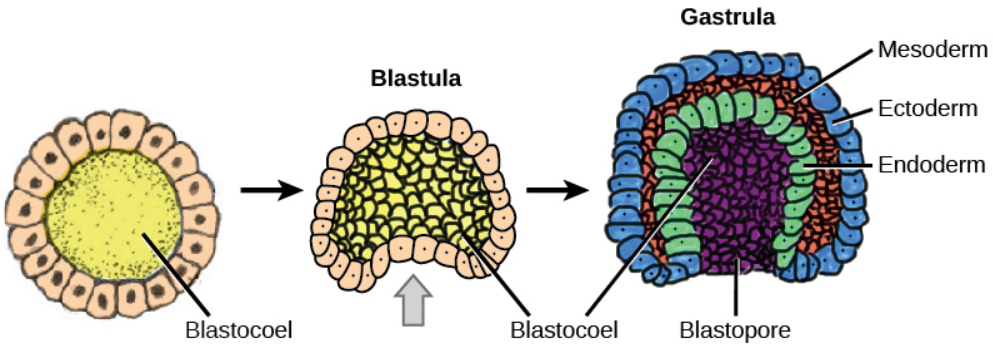


Figure 13.10 Gastrulation is the process wherein the cells in the blastula rearrange themselves to form the germ layers. (credit: modification of work by Abigail Pyne)

The three germ layers are the endoderm, the ectoderm, and the mesoderm. Cells in each germ layer differentiate into tissues and embryonic organs. The ectoderm gives rise to the nervous system and the epidermis, among other tissues. The mesoderm gives rise to the muscle cells and connective tissue in the body. The endoderm gives rise to the gut and many internal organs.

Organogenesis

Gastrulation leads to the formation of the three germ layers that give rise during further development to the different organs in the animal body. This process is called organogenesis.

Organs develop from the germ layers through the process of differentiation. During differentiation, the

embryonic stem cells express specific sets of genes that will determine their ultimate cell type. For example, some cells in the ectoderm will express the genes specific to skin cells. As a result, these cells will take on the shape and characteristics of epidermal cells. The process of differentiation is regulated by location-specific chemical signals from the cell's embryonic environment that sets in play a cascade of events that regulates gene expression.

Section Summary

The early stages of embryonic development begin with fertilization. The process of fertilization is tightly controlled to ensure that only one sperm fuses with one egg. After fertilization, the zygote undergoes cleavage to form the blastula. The blastula, which in some species is a hollow ball of cells, undergoes a process called gastrulation, during which the three germ layers form. The ectoderm gives rise to the nervous system and the epidermal skin cells, the mesoderm gives rise to the muscle cells and connective tissue in the body, and the endoderm gives rise to the digestive system and other internal organs. Organogenesis is the formation of organs from the germ layers. Each germ layer gives rise to specific tissue types.

Exercises

1. The process of gastrulation forms the _____.
    1. blastula
    2. zygote
    3. organs
    4. germ layers
2. Which of the following gives rise to the skin cells?
    1. ectoderm
    2. endoderm
    3. mesoderm
    4. none of the above
3. What do you think would happen if multiple sperm fused with one egg?

Answers

1. D
2. A
3. If multiple sperm fused with one egg, a zygote with a multiple ploidy level (multiple copies of the chromosomes) would form, and then would die.

Glossary

blastocyst: the structure formed when cells in the mammalian blastula separate into an inner and outer layer
gastrulation: the process in which the blastula folds over itself to form the three germ layers
inner cell mass: the inner layer of cells in the blastocyst, which becomes the embryo
organogenesis: the process of organ formation during development
trophoblast: the outer layer of cells in the blastocyst, which gives rise to the embryo's contribution to the placenta
zona pellucida: the protective layer around the mammalian egg

13.3 Human Reproduction

Learning Objectives

By the end of this section, you will be able to:

- Describe human male and female reproductive anatomies
- Describe spermatogenesis and oogenesis and discuss their differences and similarities
- Describe the role of hormones in human reproduction
- Describe the roles of male and female reproductive hormones

As in all animals, the adaptations for reproduction in humans are complex. They involve specialized and different anatomies in the two sexes, a hormone regulation system, and specialized behaviors regulated by the brain and endocrine system.

Human Reproductive Anatomy

The reproductive tissues of male and female humans develop similarly *in utero* until about the seventh week of gestation when a low level of the hormone testosterone is released from the gonads of the developing male. Testosterone causes the primitive gonads to differentiate into male sexual organs. When testosterone is absent, the primitive gonads develop into ovaries. Tissues that produce a penis in males produce a clitoris in females. The tissue that will become the scrotum in a male becomes the labia in a female. Thus the male and female anatomies arise from a divergence in the development of what were once common embryonic structures.

Male Reproductive Anatomy

Sperm are immobile at body temperature; therefore, the testes are external to the body so that a correct temperature is maintained for motility. In land mammals, including humans, the pair of testes must be suspended outside the body so the environment of the sperm is about 2 °C lower than body temperature to produce viable sperm. If the testes do not descend through the abdominal cavity during fetal development, the individual has reduced fertility.

The scrotum houses the testicles or testes (singular: testis), and provides passage for blood vessels, nerves, and muscles related to testicular function. The testes are a pair of male gonads that produce sperm and reproductive hormones. Each testis is approximately 2.5 by 3.8 cm (1.5 by 1 inch) in size and divided into wedge-shaped lobes by septa. Coiled in each wedge are seminiferous tubules that produce sperm.

The penis drains urine from the urinary bladder and is a copulatory organ during intercourse (Figure 13.12; Table 13.1). The penis contains three tubes of erectile tissue that become engorged with blood, making the penis erect, in preparation for intercourse. The organ is inserted into the vagina culminating with an ejaculation. During orgasm, the accessory organs and glands connected to the testes contract and empty the semen (containing sperm) into the urethra and the fluid is expelled from the body by muscular contractions causing ejaculation. After intercourse, the blood drains from the erectile tissue and the penis becomes flaccid.

Semen is a mixture of sperm (about five percent of the total) and fluids from accessory glands that contribute most of the semen's volume. Sperm are haploid cells, consisting of a flagellum for motility, a neck that contains the cell's energy-producing mitochondria, and a head that contains the genetic material (Figure 13.11). An acrosome (acrosomal vesicle) is found at the top of the head of the sperm. This structure contains enzymes that can digest the protective coverings that surround the egg and allow the sperm to fuse with the egg. An ejaculate will contain from two to five milliliters of fluid and from 50–120 million sperm per milliliter.

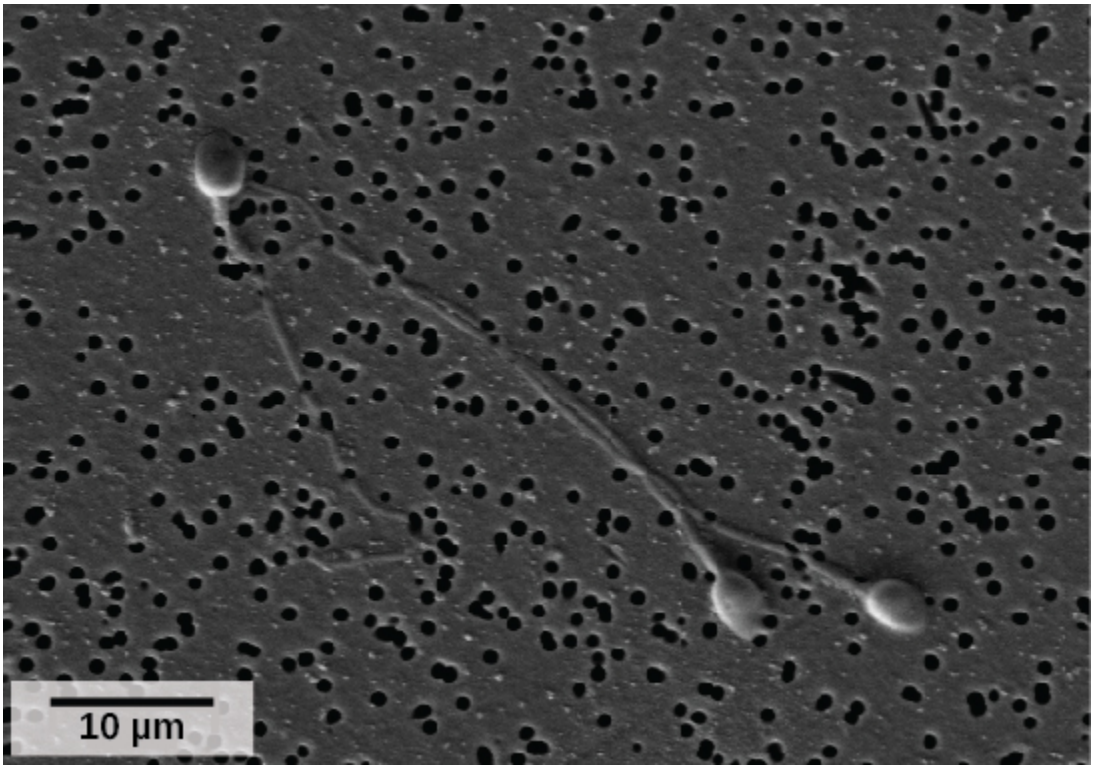


Figure 13.11 As seen in this scanning electron micrograph, human sperm has a flagellum, neck, and head. (credit: scale-bar data from Matt Russell)

Sperm form in the walls of seminiferous tubules that are coiled inside the testes (Figure 13.12; Table 13.1). The walls of the seminiferous tubules are made up of the developing sperm cells, with the least developed sperm at the periphery of the tubule and the fully developed sperm next to the lumen. The sperm cells are associated with Sertoli cells that nourish and promote the development of the sperm. Other cells present between the walls of the tubules are the interstitial cells of Leydig, which produce testosterone once the male reaches adolescence.

When the sperm have developed flagella they leave the seminiferous tubules and enter the epididymis

(Figure 13.12; Table 13.1). This structure lies along the top and posterior of the testes and is the site of sperm maturation. The sperm leave the epididymis and enter the vas deferens, which carries the sperm behind the bladder, and forms the ejaculatory duct with the duct from the seminal vesicles. During a vasectomy, a section of the vas deferens is removed, preventing sperm (but not the secretions of the accessory glands) from being passed out of the body during ejaculation and preventing fertilization.

The bulk of the semen comes from the accessory glands associated with the male reproductive system. These are the seminal vesicles, the prostate gland, and the bulbourethral gland (Figure 13.12; Table 13.1). The secretions from the accessory glands provide important compounds for the sperm including nutrients, electrolytes, and pH buffering. There are also coagulation factors that affect sperm delivery and motility.

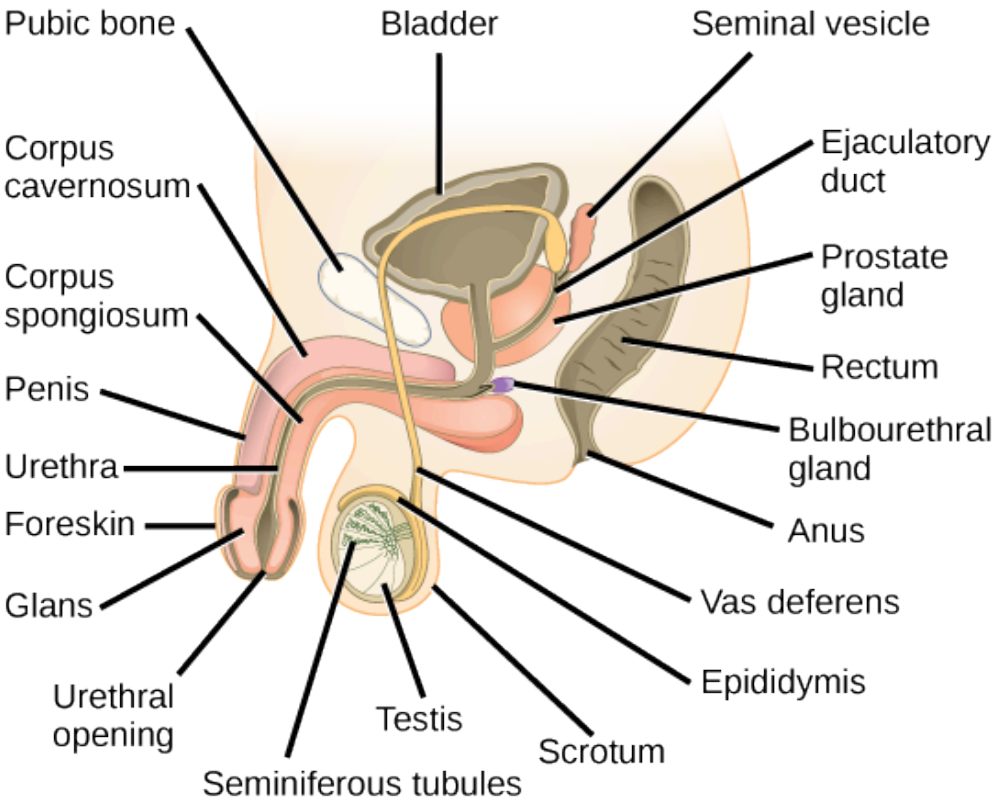


Figure 13.12 The reproductive structures of the human male are shown.

Which of the following statements about the male reproductive system is false?

A. The vas deferens carries sperm from the testes to the seminal vesicles.

B. The ejaculatory duct joins the urethra.

C. Both the prostate and the bulbourethral glands produce components of the semen.

D. The prostate gland is located in the testes.

<!–D–>

Table 13.1 Male Reproductive Anatomy

Organ	Location	Function
Scrotum	External	Supports testes and regulates their temperature
Penis	External	Delivers urine, copulating organ
Testes	Internal	Produce sperm and male hormones
Seminal Vesicles	Internal	Contribute to semen production
Prostate Gland	Internal	Contributes to semen production
Bulbourethtral Glands	Internal	Neutralize urine in urethra

Female Reproductive Anatomy

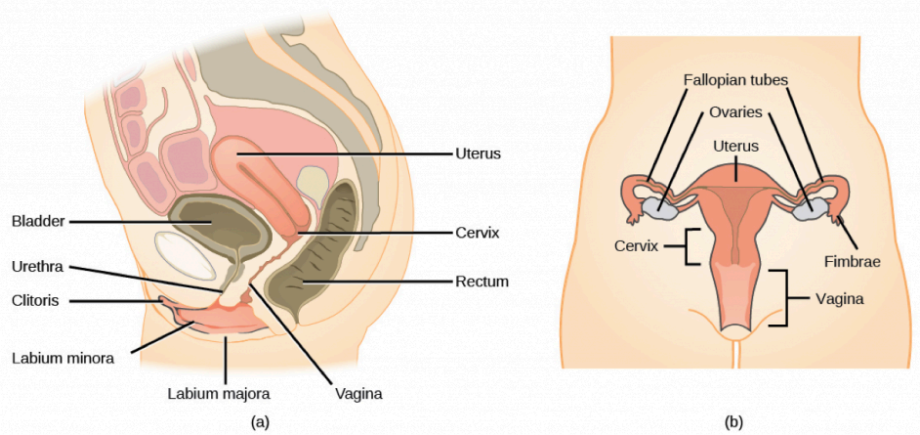


Figure 13.13 A number of female reproductive structures are exterior to the body. These include the breasts and the vulva, which consists of the mons pubis, clitoris, labia majora, labia minora, and the vestibular glands (Figure 13.13; Table 13.2).

The breasts consist of mammary glands and fat. Each gland consists of 15 to 25 lobes that have ducts that empty at the nipple and that supply the nursing child with nutrient- and antibody-rich milk to aid development and protect the child.

Internal female reproductive structures include ovaries, oviducts, the uterus, and the vagina (Figure 13.13; Table 13.2). The pair of ovaries is held in place in the abdominal cavity by a system of ligaments. The outermost layer of the ovary is made up of follicles, each consisting of one or more follicular cells that surround, nourish, and protect a single egg. During the menstrual period, a batch of follicular cells develops and prepares their eggs for release. At ovulation, one follicle ruptures and one egg is released. Following ovulation, the follicular tissue that surrounded the ovulated egg stays within the ovary and grows to form a solid mass called the corpus luteum. The corpus luteum secretes additional estrogen and the hormone progesterone that helps maintain the uterine lining during pregnancy. The ovaries also produce hormones, such as estrogen.

The oviducts, or fallopian tubes, extend from the uterus in the lower abdominal cavity to the ovaries, but they are not in contact with the ovaries. The lateral ends of the oviducts flare out into a trumpet-like

structure and have a fringe of finger-like projections called fimbrae. When an egg is released at ovulation, the fimbrae help the nonmotile egg enter into the tube. The walls of the oviducts have a ciliated epithelium over smooth muscle. The cilia beat, and the smooth muscle contracts, moving the egg toward the uterus. Fertilization usually takes place within the oviduct and the developing embryo is moved toward the uterus. It usually takes the egg or embryo a week to travel through the oviduct.

Sterilization in women is called a tubal ligation; it is analogous to a vasectomy in males in that the oviducts are severed and sealed, preventing sperm from reaching the egg.

The uterus is a structure about the size of a woman's fist. The uterus has a thick muscular wall and is lined with an endometrium rich in blood vessels and mucus glands that develop and thicken during the female cycle. Thickening of the endometrium prepares the uterus to receive the fertilized egg or zygote, which will then implant itself in the endometrium. The uterus supports the developing embryo and fetus during gestation. Contractions of the smooth muscle in the uterus aid in forcing the baby through the vagina during labor. If fertilization does not occur, a portion of the lining of the uterus sloughs off during each menstrual period. The endometrium builds up again in preparation for implantation. Part of the uterus, called the cervix, protrudes into the top of the vagina.

The vagina is a muscular tube that serves several purposes. It allows menstrual flow to leave the body. It is the receptacle for the penis during intercourse and the pathway for the delivery of offspring.

Table 13.2 Female Reproductive Anatomy

Organ	Location	Function
Clitoris	External	Sensory organ
Mons pubis	External	Fatty area overlying pubic bone
Labia majora	External	Covers labia minora; contains sweat and sebaceous glands
Labia minora	External	Covers vestibule
Greater vestibular glands	External	Secrete mucus; lubricate vagina
Breast	External	Produces and delivers milk
Ovaries	Internal	Produce and develop eggs
Oviducts	Internal	Transport egg to uterus; site of fertilization
Uterus	Internal	Supports developing embryo
Vagina	Internal	Common tube for intercourse, birth canal, passing menstrual flow

Gametogenesis (Spermatogenesis and Oogenesis)

Gametogenesis, the production of sperm and eggs, involves the process of meiosis. During meiosis, two nuclear divisions separate the paired chromosomes in the nucleus and then separate the chromatids that were made during an earlier stage of the cell's life cycle. Meiosis and its associated cell divisions

produces haploid cells with half of each pair of chromosomes normally found in diploid cells. The production of sperm is called spermatogenesis and the production of eggs is called oogenesis.

Spermatogenesis

Spermatogenesis occurs in the wall of the seminiferous tubules, with the most primitive cells at the periphery of the tube and the most mature sperm at the lumen of the tube (Figure 13.14). Immediately under the capsule of the tubule are diploid, undifferentiated cells. These stem cells, each called a spermatogonium (pl. spermatogonia), go through mitosis to produce one cell that remains as a stem cell and a second cell called a primary spermatocyte that will undergo meiosis to produce sperm.

The diploid primary spermatocyte goes through meiosis I to produce two haploid cells called secondary spermatocytes. Each secondary spermatocyte divides after meiosis II to produce two cells called spermatids. The spermatids eventually reach the lumen of the tubule and grow a flagellum, becoming sperm cells. Four sperm result from each primary spermatocyte that goes through meiosis.

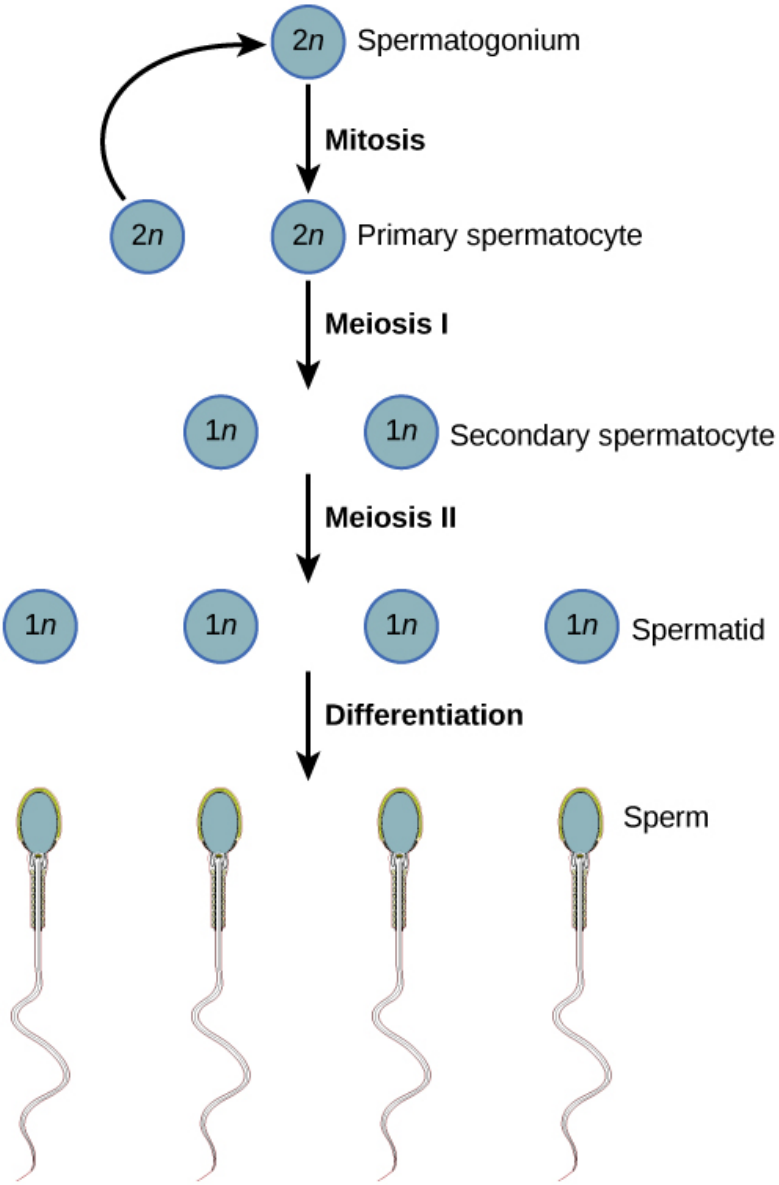


Figure 13.14 During spermatogenesis, four sperm result from each primary spermatocyte. The process also maps onto the physical structure of the wall of the seminiferous tubule, with the spermatogonia on the outer side of the tubule, and the sperm with their developing tails extended into the lumen of the tubule.

Concept in Action

Visit this site to see the process of spermatogenesis.

Oogenesis

Oogenesis occurs in the outermost layers of the ovaries. As with sperm production, oogenesis starts with a germ cell. In oogenesis, this germ cell is called an oogonium and forms during the embryological development of the individual. The oogonium undergoes mitosis to produce about one to two million oocytes by the time of birth.

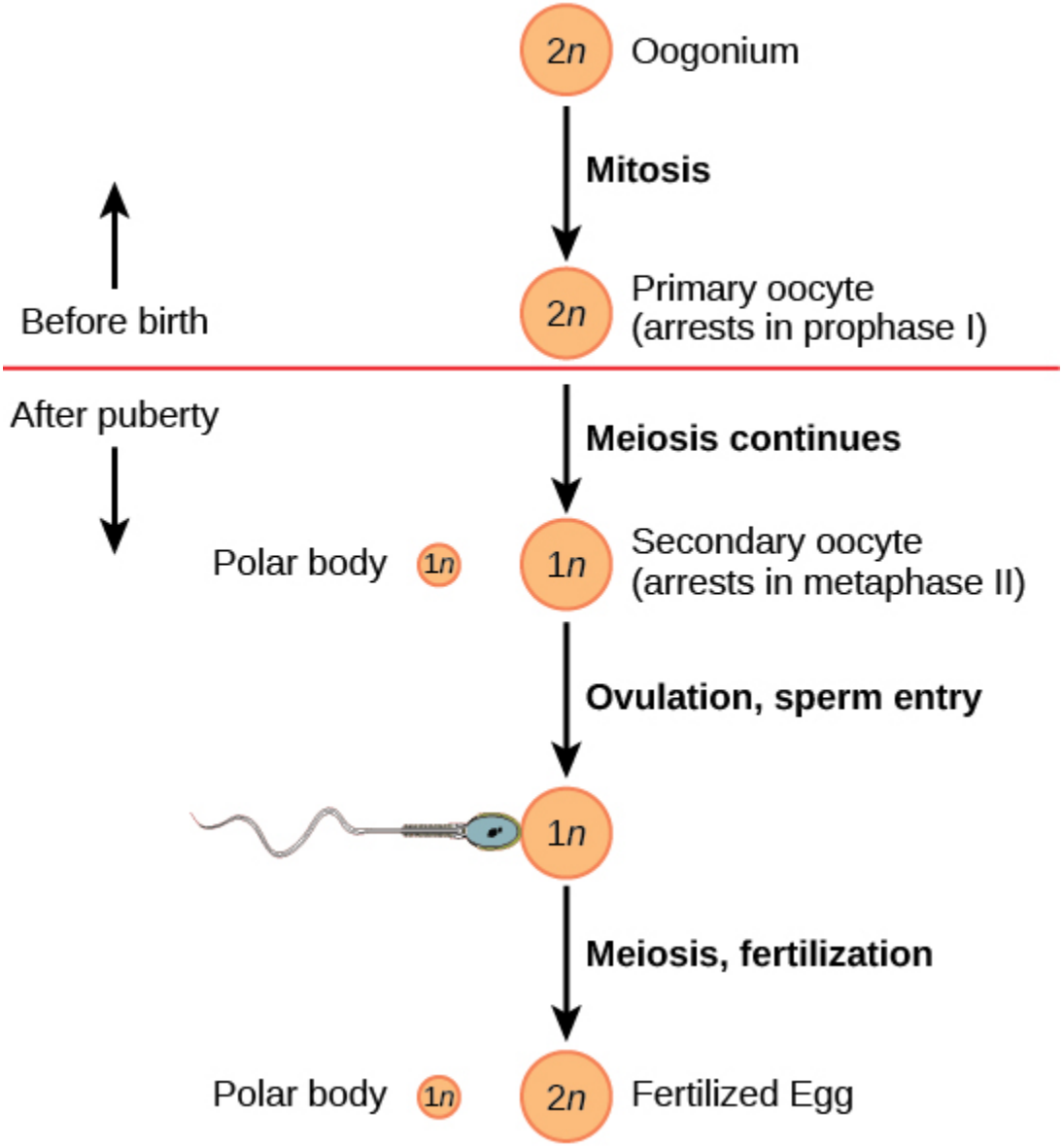


Figure 13.15 The process of oogenesis occurs in the ovary's outermost layer.

The primary oocytes begin meiosis before birth (Figure 13.15). However, the meiotic division is arrested in its progress in the first prophase stage. At the time of birth, all future eggs are in prophase I. This situation is in contrast with the male reproductive system in which sperm are produced continuously throughout the life of the individual. Starting at adolescence, anterior pituitary hormones cause the development of a few follicles in an ovary each month. This results in a primary oocyte finishing the first meiotic division. The cell divides unequally, with most of the cytoplasm and organelles going to one cell, called a secondary oocyte, and only one set of chromosomes and a small amount of cytoplasm going to the other cell. This second cell is called a polar body and usually dies. Cell division is again arrested, this time at metaphase II. At ovulation, this secondary oocyte is released and travels toward the uterus through the oviduct. If the secondary oocyte is fertilized, the cell continues through meiosis II, producing a second polar body and haploid egg, which fuses with the haploid sperm to form a fertilized egg (zygote) containing all 46 chromosomes.

Hormonal Control of Reproduction

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 13.16). 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.

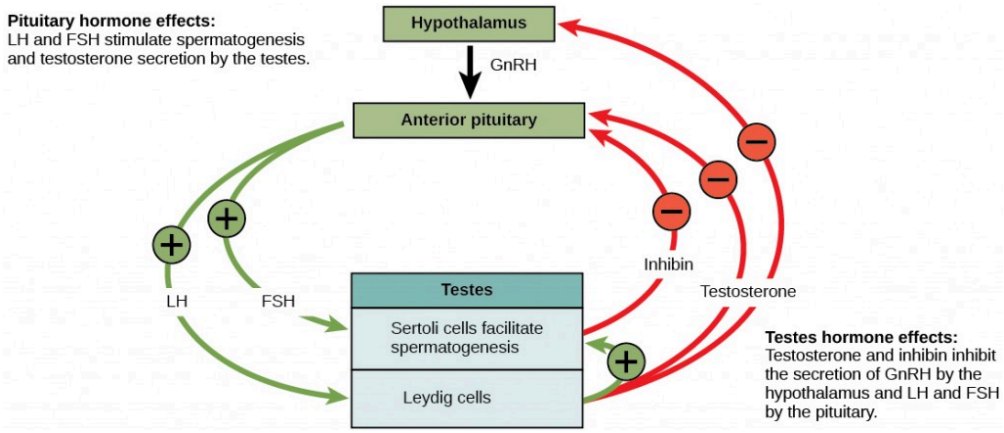


Figure 13.16 Hormones control sperm production in a negative feedback system.

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 13.17). 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 13.17). 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.

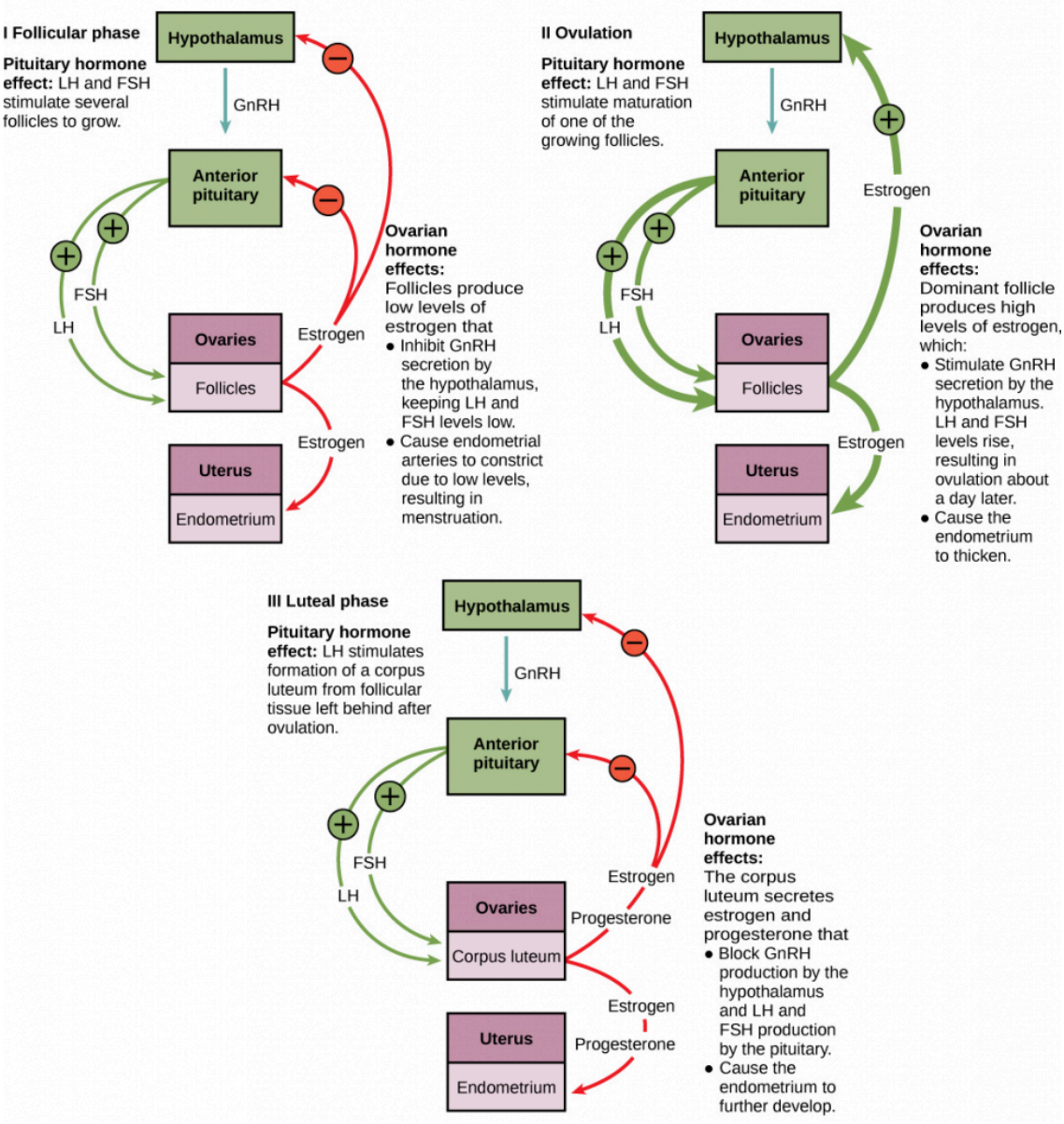


Figure 13.17 The ovarian and menstrual cycles of female reproduction are regulated by hormones produced by the hypothalamus, pituitary, and ovaries.

Which of the following statements about hormone regulation of the female reproductive cycle is false?

A. LH and FSH are produced in the pituitary, and estrogen and progesterone are produced in the ovaries.

B. Estradiol and progesterone secreted from the corpus luteum cause the endometrium to thicken.

C. Both progesterone and estrogen are produced by the follicles.

D. Secretion of GnRH by the hypothalamus is inhibited by low levels of estrogen but stimulated by high levels of estrogen.

<!– C–>

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.

Reproductive Endocrinologist

A reproductive endocrinologist is a physician who treats a variety of hormonal disorders related to reproduction and infertility in both men and women. The disorders include menstrual problems, infertility, pregnancy loss, sexual dysfunction, and menopause. Doctors may use fertility drugs, surgery, or assisted reproductive techniques (ART) in their therapy. ART involves the use of procedures to manipulate the egg or sperm to facilitate reproduction, such as *in vitro* fertilization.

Reproductive endocrinologists undergo extensive medical training, first in a four-year residency in obstetrics and gynecology, then in a three-year fellowship in reproductive endocrinology. To be board certified in this area, the physician must pass written and oral exams in both areas.

Gestation

Pregnancy begins with the fertilization 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.

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 13.18 **a**). 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.

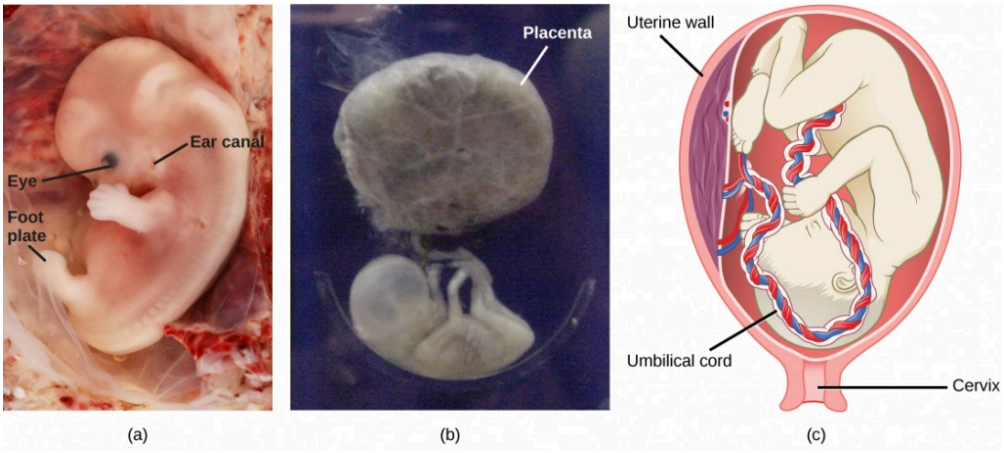


Figure 13.18 (a) Fetal development is shown at nine weeks gestation. (b) This fetus is just entering the second trimester, when the placenta takes over more of the functions performed as the baby develops. (c) There is rapid fetal growth during the third trimester. (credit a: modification of work by 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 13.18 **b**). 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 13.18 **c**). This is the period of the most rapid growth during the pregnancy as all organ systems continue to grow and develop.

Concept in Action

Visit this website to see the stages of human fetal development.

Labor is the muscular contractions to 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 baby reorients, facing forward and down with the back or crown of the head engaging 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.

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.

Section Summary

The reproductive structures that evolved in land animals allow males and females to mate, fertilize internally, and support the growth and development of offspring. Gametogenesis, the production of sperm (spermatogenesis) and eggs (oogenesis), takes place through the process of meiosis.

The male and female reproductive cycles are controlled by hormones released from the hypothalamus and anterior pituitary and hormones from reproductive tissues and organs. The hypothalamus monitors the need for FSH and LH production and release from the anterior pituitary. FSH and LH affect reproductive structures to cause the formation of sperm and the preparation of eggs for release and possible fertilization. In the male, FSH and LH stimulate Sertoli cells and interstitial cells of Leydig in the testes to facilitate sperm production. The Leydig cells produce testosterone, which also is responsible for the secondary sexual characteristics of males. In females, FSH and LH cause estrogen and progesterone to be produced. They regulate the female reproductive cycle, which is divided into the ovarian cycle and the menstrual cycle.

Human pregnancy begins with fertilization of an egg and proceeds through the three trimesters of

gestation. The first trimester lays down the basic structures of the body, including the limb buds, heart, eyes, and the liver. The second trimester continues the development of all of the organs and systems. The third trimester exhibits the greatest growth of the fetus and culminates in labor and delivery. The labor process has three stages (contractions, delivery of the fetus, and expulsion of the placenta), each propelled by hormones.

Exercises

1. Which of the following statements about the male reproductive system is false?
    1. The vas deferens carries sperm from the testes to the seminal vesicles.
    2. The ejaculatory duct joins the urethra.
    3. Both the prostate and the bulbourethral glands produce components of the semen.
    4. The prostate gland is located in the testes.
2. Which of the following statements about hormone regulation of the female reproductive cycle is false?
    1. LH and FSH are produced in the pituitary, and estrogen and progesterone are produced in the ovaries.
    2. Estradiol and progesterone secreted from the corpus luteum cause the endometrium to thicken.
    3. Both progesterone and estrogen are produced by the follicles.
    4. Secretion of GnRH by the hypothalamus is inhibited by low levels of estrogen but stimulated by high levels of estrogen.
3. Sperm are produced in the _____.
    1. scrotum
    2. seminal vesicles
    3. seminiferous tubules
    4. prostate gland
4. Which female organ has an endometrial lining that will support a developing baby?
    1. labia minora
    2. breast
    3. ovaries
    4. uterus
5. Which hormone causes FSH and LH to be released?
    1. testosterone
    2. estrogen
    3. GnRH

4. progesterone

6. Nutrient and waste requirements for the developing fetus are handled during the first few weeks by _____.

    1. the placenta
    2. diffusion through the endometrium
    3. the chorion
    4. the blastocyst

7. Which hormone is primarily responsible for the contractions during labor?

    1. oxytocin
    2. estrogen
    3. β-HCG
    4. progesterone

8. Compare spermatogenesis and oogenesis as to timing of the processes, and the number and type of cells finally produced.
9. Describe the events in the ovarian cycle leading up to ovulation.
10. Describe the stages of labor.

Answers

1. D
2. C
3. C
4. D
5. C
6. B
7. A
8. Stem cells are laid down in the male during gestation and lie dormant until adolescence. Stem cells in the female increase to one to two million and enter the first meiotic division and are arrested in prophase. At adolescence, spermatogenesis begins and continues until death, producing the maximum number of sperm with each meiotic division. Oogenesis continues again at adolescence in batches of eggs with each menstrual cycle. These primary oocytes finish the first meiotic division, producing a viable egg with most of the cytoplasm and its contents, and a second cell called a polar body containing 23 chromosomes. The second meiotic division is initiated and arrested in metaphase. At ovulation, one egg is released. If this egg is fertilized, it finishes the second meiotic division. This is a diploid, fertilized egg.
9. Low levels of progesterone allow the hypothalamus to send GnRH to the anterior pituitary and cause the release of FSH and LH. FSH stimulates follicles on the ovary to grow and prepare the eggs for ovulation. As the follicles increase in size, they begin to release estrogen and a low level of progesterone into the blood. The level of estrogen rises to a peak, causing a spike in the

concentration of LH. This causes the most mature follicle to rupture and ovulation occurs.

10. Stage one of labor results in uterine contractions, which thin the cervix and dilate the cervical opening. Stage two delivers the baby, and stage three delivers the placenta.

Glossary

bulbourethral gland: the paired glands in the human male that produce a secretion that cleanses the urethra prior to ejaculation

corpus luteum: the endocrine tissue that develops from an ovarian follicle after ovulation; secretes progesterone and estrogen during pregnancy

clitoris: a sensory and erectile structure in female mammals, homologous to the male penis, stimulated during sexual arousal

estrogen: a reproductive hormone in females that assists in endometrial regrowth, ovulation, and calcium absorption

follicle stimulating hormone (FSH): a reproductive hormone that causes sperm production in men and follicle development in women

gestation: the development before birth of a viviparous animal

gestation period: the length of time of development, from conception to birth, of the young of a viviparous animal

gonadotropin-releasing hormone (GnRH): a hormone from the hypothalamus that causes the release of FSH and LH from the anterior pituitary

human beta chorionic gonadotropin (β-HCG): a hormone produced by the chorion of the zygote that helps to maintain the corpus luteum and elevated levels of progesterone

inhibin: a hormone made by Sertoli cells, provides negative feedback to hypothalamus in control of FSH and GnRH release

interstitial cell of Leydig: a cell type found next to the seminiferous tubules that makes testosterone

labia majora: the large folds of tissue covering inguinal area

labia minora: the smaller folds of tissue within labia majora

luteinizing hormone (LH): a reproductive hormone in both men and women, causes testosterone production in men and ovulation and lactation in women

menstrual cycle: the cycle of the degradation and re-growth of the endometrium

oogenesis: the process of producing haploid eggs

ovarian cycle: the cycle of preparation of egg for ovulation and the conversion of the follicle to the corpus luteum

oviduct: (also, fallopian tube) the muscular tube connecting uterus with ovary area

ovulation: the release of an oocyte from a mature follicle in the ovary of a vertebrate

penis: the male reproductive structure for urine elimination and copulation

placenta: the organ that supports the transport of nutrients and waste between the mothers and fetus' blood in eutherian mammals

progesterone: a reproductive hormone in women; assists in endometrial regrowth and inhibition of FSH and LH release

prostate gland: a structure that is a mixture of smooth muscle and glandular material and that contributes to semen

scrotum: a sac containing testes, exterior to body

semen: a fluid mixture of sperm and supporting materials

seminal vesicle: a secretory accessory gland in male; contributes to semen

seminiferous tubule: the structures within which sperm production occurs in the testes

Sertoli cell: a cell in the walls of the seminiferous tubules that assists developing sperm and secretes inhibin

spermatogenesis: the process of producing haploid sperm

testes: a pair of male reproductive organs

testosterone: a reproductive hormone in men that assists in sperm production and promoting secondary sexual characteristics

uterus: a female reproductive structure in which an embryo develops

vagina: a muscular tube for the passage of menstrual flow, copulation, and birth of offspring

Chapter 14. The Animal Body: Basic Form and Function

Figure 14.1 An arctic fox is a complex animal, well adapted to its environment. It changes coat color with the seasons, and has longer fur in winter to trap heat. (credit: modification of work by Keith Morehouse, USFWS)

The arctic fox is an example of a complex animal that has adapted to its environment and illustrates the relationships between an animal's form and function. The structures of animals consist of primary tissues that make up more complex organs and organ systems. Homeostasis allows an animal to maintain a balance between its internal and external environments.

14.1 Animal Form and Function

Learning Objectives

By the end of this section, you will be able to:

- Describe the various types of body plans that occur in animals

Animals vary in form and function. From a sponge to a worm to a goat, an organism has a distinct body plan that limits its size and shape. Animals' bodies are also designed to interact with their environments, whether in the deep sea, a rainforest canopy, or the desert. Therefore, a large amount of information about the structure of an organism's body (anatomy) and the function of its cells, tissues and organs (physiology) can be learned by studying that organism's environment.

Body Plans

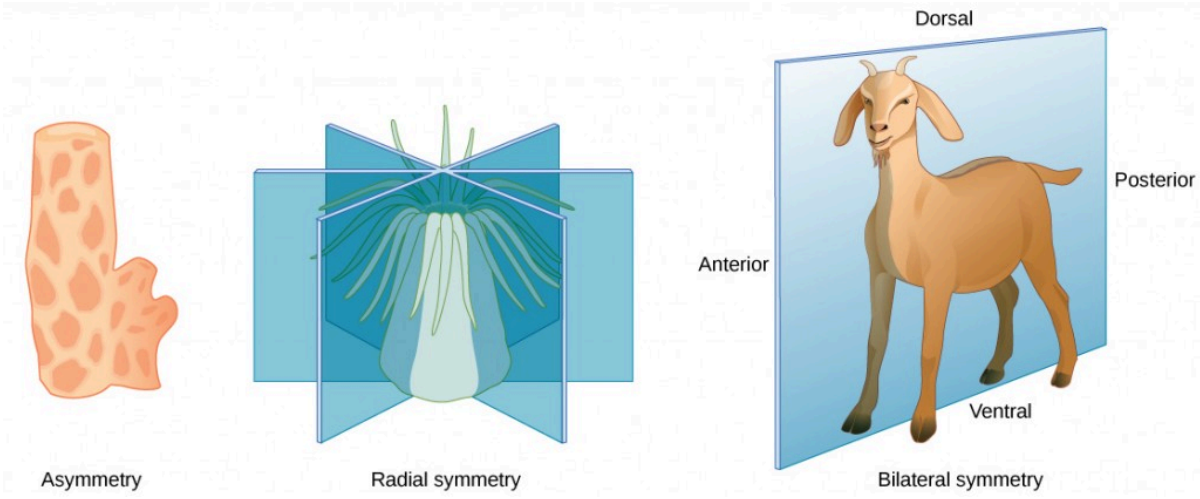


Figure 14.2 Animals exhibit different types of body symmetry. The sponge is asymmetrical, the sea anemone has radial symmetry, and the goat has bilateral symmetry.

Animal body plans follow set patterns related to symmetry. They are asymmetrical, radial, or bilateral in form as illustrated in Figure 14.2. *Asymmetrical* animals are animals with no pattern or symmetry; an example of an asymmetrical animal is a sponge. Radial symmetry, as illustrated in Figure 14.2, describes when an animal has an up-and-down orientation: any plane cut along its longitudinal axis through the organism produces equal halves, but not a definite right or left side. This plan is found mostly in aquatic animals, especially organisms that attach themselves to a base, like a rock or a boat, and extract their food from the surrounding water as it flows around the organism. Bilateral symmetry is illustrated in the same figure by a goat. The goat also has an upper and lower component to it, but a plane cut from front to back separates the animal into definite right and left sides. Additional terms

used when describing positions in the body are anterior (front), posterior (rear), dorsal (toward the back), and ventral (toward the stomach). Bilateral symmetry is found in both land-based and aquatic animals; it enables a high level of mobility.

Limiting Effects of Diffusion on Size and Development

The exchange of nutrients and wastes between a cell and its watery environment occurs through the process of diffusion. All living cells are bathed in liquid, whether they are in a single-celled organism or a multicellular one. Diffusion is effective over a specific distance and limits the size that an individual cell can attain. If a cell is a single-celled microorganism, such as an amoeba, it can satisfy all of its nutrient and waste needs through diffusion. If the cell is too large, then diffusion is ineffective and the center of the cell does not receive adequate nutrients nor is it able to effectively dispel its waste.

An important concept in understanding how efficient diffusion is as a means of transport is the surface to volume ratio. Recall that any three-dimensional object has a surface area and volume; the ratio of these two quantities is the surface-to-volume ratio. Consider a cell shaped like a perfect sphere: it has a surface area of $4\pi r^2$, and a volume of $(4/3)\pi r^3$. The surface-to-volume ratio of a sphere is $3/r$; as the cell gets bigger, its surface to volume ratio decreases, making diffusion less efficient. The larger the size of the sphere, or animal, the less surface area for diffusion it possesses.

The solution to producing larger organisms is for them to become multicellular. Specialization occurs in complex organisms, allowing cells to become more efficient at doing fewer tasks. For example, circulatory systems bring nutrients and remove waste, while respiratory systems provide oxygen for the cells and remove carbon dioxide from them. Other organ systems have developed further specialization of cells and tissues and efficiently control body functions. Moreover, surface-to-volume ratio applies to other areas of animal development, such as the relationship between muscle mass and cross-sectional surface area in supporting skeletons, and in the relationship between muscle mass and the generation of dissipation of heat.

Concept in Action

Visit this interactive site to see an entire animal (a zebrafish embryo) at the cellular and sub-cellular level. Use the zoom and navigation functions for a virtual nanoscopy exploration.

Animal Bioenergetics

All animals must obtain their energy from food they ingest or absorb. These nutrients are converted to adenosine triphosphate (ATP) for short-term storage and use by all cells. Some animals store energy for

slightly longer times as glycogen, and others store energy for much longer times in the form of triglycerides housed in specialized adipose tissues. No energy system is one hundred percent efficient, and an animal's metabolism produces waste energy in the form of heat. If an animal can conserve that heat and maintain a relatively constant body temperature, it is classified as a warm-blooded animal and called an **endotherm**. The insulation used to conserve the body heat comes in the forms of fur, fat, or feathers. The absence of insulation in **ectothermic** animals increases their dependence on the environment for body heat.

The amount of energy expended by an animal over a specific time is called its metabolic rate. The rate is measured variously in joules, calories, or kilocalories (1000 calories). Carbohydrates and proteins contain about 4.5 to 5 kcal/g, and fat contains about 9 kcal/g. Metabolic rate is estimated as the **basal metabolic rate (BMR)** in endothermic animals at rest and as the **standard metabolic rate (SMR)** in ectotherms. Human males have a BMR of 1600 to 1800 kcal/day, and human females have a BMR of 1300 to 1500 kcal/day. Even with insulation, endothermal animals require extensive amounts of energy to maintain a constant body temperature. An ectotherm such as an alligator has an SMR of 60 kcal/day.

Energy Requirements Related to Body Size

Smaller endothermic animals have a greater surface area for their mass than larger ones (Figure 14.4). Therefore, smaller animals lose heat at a faster rate than larger animals and require more energy to maintain a constant internal temperature. This results in a smaller endothermic animal having a higher BMR, per body weight, than a larger endothermic animal.

Species		
Mass	35 g	4,500,000 g
Metabolic rate	890 mm^3 0_2/g body mass/hr	75 mm^3 0_2/g body mass/hr

Figure 14.4. The mouse has a much higher metabolic rate than the elephant. (credit "mouse": modification of work by Magnus Kjaergaard; credit "elephant": modification of work by "TheLizardQueen"/Flickr)

Energy Requirements Related to Levels of Activity

The more active an animal is, the more energy is needed to maintain that activity, and the higher its BMR or SMR. The average daily rate of energy consumption is about two to four times an animal's BMR or SMR. Humans are more sedentary than most animals and have an average daily rate of only 1.5 times the BMR. The diet of an endothermic animal is determined by its BMR. For example: the type of grasses, leaves, or shrubs that an herbivore eats affects the number of calories that it takes in.

The relative caloric content of herbivore foods, in descending order, is tall grasses > legumes > short grasses > forbs (any broad-leaved plant, not a grass) > subshrubs > annuals/biennials.

Energy Requirements Related to Environment

Animals adapt to extremes of temperature or food availability through torpor. **Torpor** is a process that leads to a decrease in activity and metabolism and allows animals to survive adverse conditions. Torpor can be used by animals for long periods, such as entering a state of **hibernation** during the winter months, in which case it enables them to maintain a reduced body temperature. During hibernation, ground squirrels can achieve an abdominal temperature of 0° C (32° F), while a bear's internal temperature is maintained higher at about 37° C (99° F).

If torpor occurs during the summer months with high temperatures and little water, it is called **estivation**. Some desert animals use this to survive the harshest months of the year. Torpor can occur on a daily basis; this is seen in bats and hummingbirds. While endothermy is limited in smaller animals by surface to volume ratio, some organisms can be smaller and still be endotherms because they employ daily torpor during the part of the day that is coldest. This allows them to conserve energy during the colder parts of the day, when they consume more energy to maintain their body temperature.

Animal Body Planes and Cavities

A standing vertebrate animal can be divided by several planes. A **sagittal plane** divides the body into right and left portions. A **midsagittal plane** divides the body exactly in the middle, making two equal right and left halves. A **frontal plane** (also called a coronal plane) separates the front from the back. A **transverse plane** (or, horizontal plane) divides the animal into upper and lower portions. This is sometimes called a cross section, and, if the transverse cut is at an angle, it is called an oblique plane. Figure 14.5 illustrates these planes on a goat (a four-legged animal) and a human being.

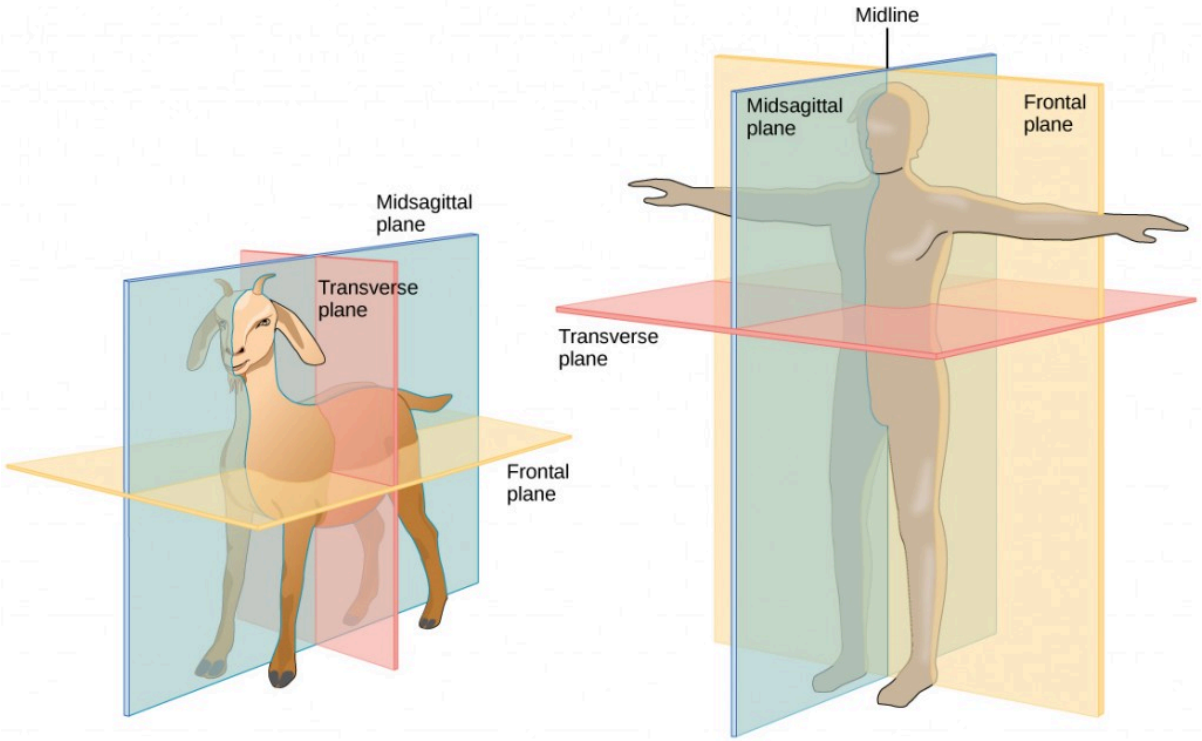


Figure 14.5. Shown are the planes of a quadruped goat and a bipedal human. The midsagittal plane divides the body exactly in half, into right and left portions. The frontal plane divides the front and back, and the transverse plane divides the body into upper and lower portions.

Vertebrate animals have a number of defined body cavities, as illustrated in Figure 14.6. Two of these are major cavities that contain smaller cavities within them. The **dorsal cavity** contains the cranial and the vertebral (or spinal) cavities. The **ventral cavity** contains the thoracic cavity, which in turn contains the pleural cavity around the lungs and the pericardial cavity, which surrounds the heart. The ventral cavity also contains the abdominopelvic cavity, which can be separated into the abdominal and the pelvic cavities.

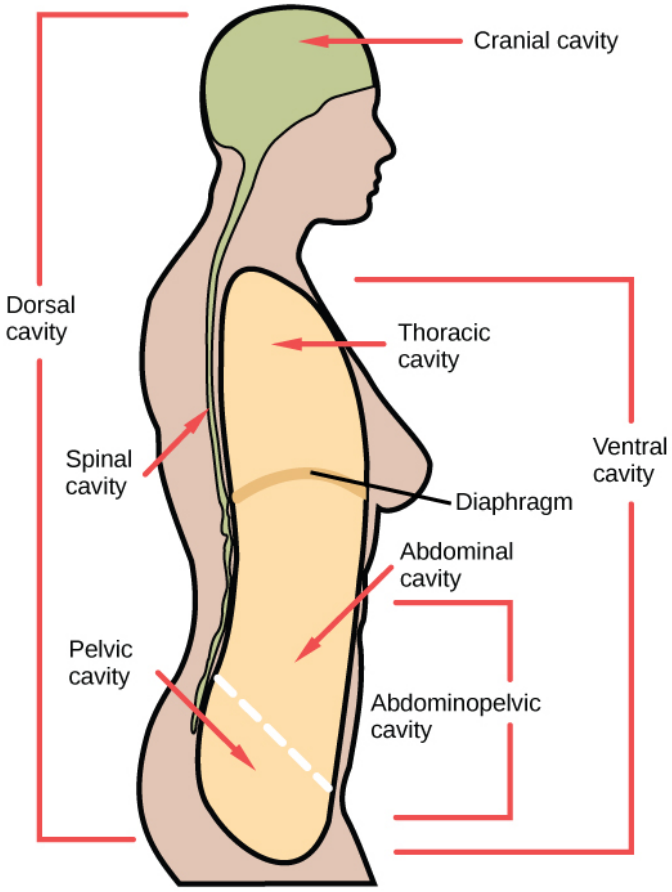


Figure 14.6. Vertebrate animals have two major body cavities. The dorsal cavity, indicated in green, contains the cranial and the spinal cavity. The ventral cavity, indicated in yellow, contains the thoracic cavity and the abdominopelvic cavity. The thoracic cavity is separated from the abdominopelvic cavity by the diaphragm. The thoracic cavity is separated into the abdominal cavity and the pelvic cavity by an imaginary line parallel to the pelvis bones. (credit: modification of work by NCI)

Physical Anthropologist

Physical anthropologists study the adaption, variability, and evolution of human beings, plus their living and fossil relatives. They can work in a variety of settings, although most will have an academic appointment at a university, usually in an anthropology department or a biology, genetics, or zoology department.

Non-academic positions are available in the automotive and aerospace industries where the focus is on human size, shape, and anatomy. Research by these professionals might range from studies of how the human body reacts to car crashes to exploring how to make seats more comfortable. Other non-academic positions can be obtained in museums of natural history, anthropology, archaeology, or science and technology. These positions involve educating students from grade school through graduate school. Physical anthropologists serve as education coordinators, collection managers, writers for museum publications, and as administrators. Zoos employ these professionals, especially if they have

an expertise in primate biology; they work in collection management and captive breeding programs for endangered species. Forensic science utilizes physical anthropology expertise in identifying human and animal remains, assisting in determining the cause of death, and for expert testimony in trials.

Glossary

asymmetrical: describes animals with no axis of symmetry in their body pattern
basal metabolic rate (BMR): metabolic rate at rest in endothermic animals
cartilage: type of connective tissue with a large amount of ground substance matrix, cells called chondrocytes, and some amount of fibers
chondrocyte: cell found in cartilage
columnar epithelia: epithelia made of cells taller than they are wide, specialized in absorption
connective tissue: type of tissue made of cells, ground substance matrix, and fibers
cuboidal epithelia: epithelia made of cube-shaped cells, specialized in glandular functions
dorsal cavity: body cavity on the posterior or back portion of an animal; includes the cranial and vertebral cavities
ectotherm: animal incapable of maintaining a relatively constant internal body temperature**endotherm:** animal capable of maintaining a relatively constant internal body temperature
estivation: torpor in response to extremely high temperatures and low water availability
fibrous connective tissue: type of connective tissue with a high concentration of fibers
fusiform: animal body shape that is tubular and tapered at both ends
hibernation: torpor over a long period of time, such as a winter
homeostasis: dynamic equilibrium maintaining appropriate body functions
lacuna: space in cartilage and bone that contains living cells
matrix: component of connective tissue made of both living and non-living (ground substances) cells
midsagittal plane: plane cutting through an animal separating the individual into even right and left sides
negative feedback loop: feedback to a control mechanism that increases or decreases a stimulus instead of maintaining it
osteon: subunit of compact bone
positive feedback loop:feedback to a control mechanism that continues the direction of a stimulus
sagittal plane: plane cutting through an animal separating the individual into right and left sides
set point: midpoint or target point in homeostasis
squamous epithelia: type of epithelia made of flat cells, specialized in aiding diffusion or preventing abrasion
standard metabolic rate (SMR): metabolic rate at rest in ectothermic animals
stratified epithelia: multiple layers of epithelial cells
torpor: decrease in activity and metabolism that allows an animal to survive adverse conditions
ventral cavity: body cavity on the anterior or front portion of an animal that includes the thoracic cavities and the abdominopelvic cavities

14.2 Animal Primary Tissues

Learning Objectives

By the end of this section, you will be able to:

- Describe epithelial tissues
- Discuss the different types of connective tissues in animals
- Describe three types of muscle tissues
- Describe nervous tissue

Multicellular, complex animals have four primary types of tissue: epithelial, connective, muscle, and nervous. Recall that tissues are groups of similar cells carrying out related functions. These tissues combine to form organs—like the skin or kidney—that have specific, 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 are called *simple* **epithelia**; epithelial tissue composed of multiple layers is called **stratified epithelia**. Table 14.2 summarizes the different types of epithelial tissues.

Table 14.2 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

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.

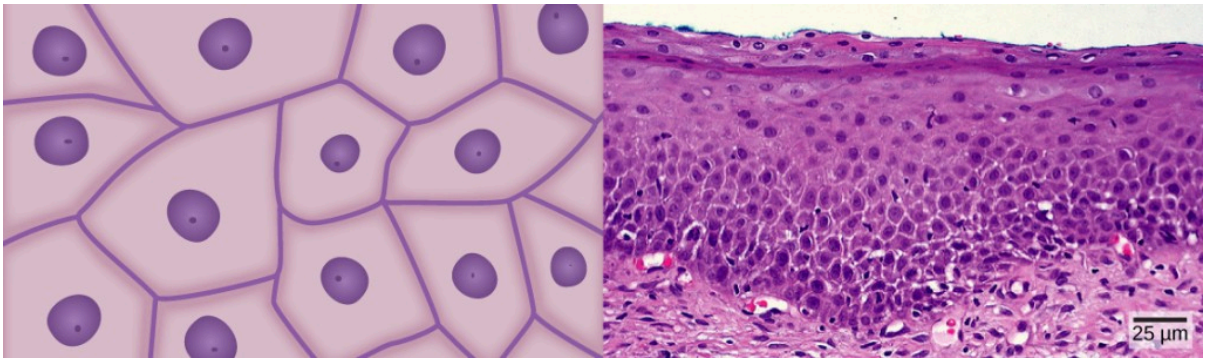


Figure 14.7 Squamous epithelia cells (a) have a slightly irregular shape, and a small, centrally located nucleus. These cells can be stratified into layers, as in (b) this human cervix specimen. (credit b: modification of work by Ed Uthman; scale-bar data from Matt Russell)

Figure 14.7**a** illustrates a layer of squamous cells with their membranes joined together to form an epithelium. Image Figure 14.7**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 14.8, are cube-shaped with a single, central nucleus. They are most commonly found in a single layer representing a simple epithelia in glandular tissues throughout the body where they prepare and secrete glandular material. They are also found in the walls of tubules and in the ducts of the kidney and liver.

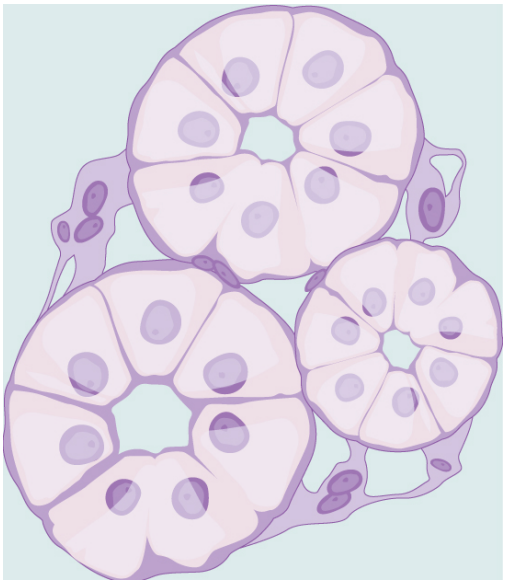


Figure 14.8. Simple cuboidal epithelial cells line tubules in the mammalian kidney, where they are involved in filtering the blood.

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 14.9. 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.

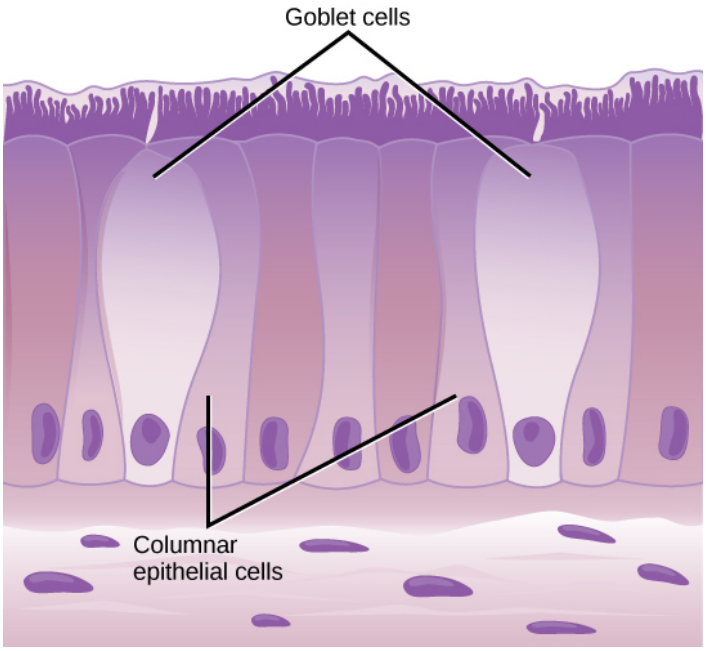


Figure 14.9. Simple columnar epithelial cells absorb material from the digestive tract. Goblet cells secret mucous into the digestive tract lumen.

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 14.10. 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.

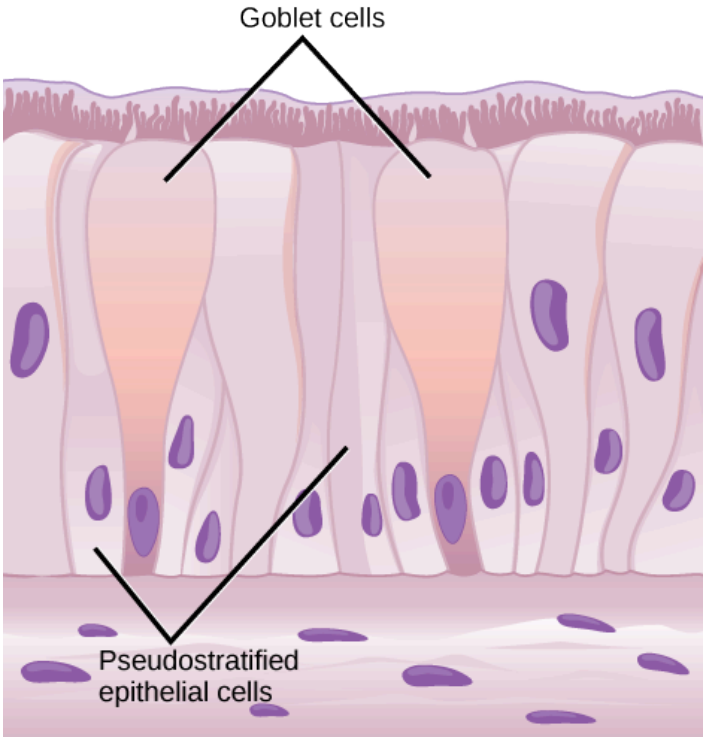


Figure 14.10. Pseudostratified columnar epithelia line the respiratory tract. They exist in one layer, but the arrangement of nuclei at different levels makes it appear that there is more than one layer. Goblet cells interspersed between the columnar epithelial cells secrete mucous into the respiratory tract.

Transitional Epithelia

Transitional or uroepithelial cells appear only in the urinary system, primarily in the bladder and ureter. These cells are arranged in a stratified layer, but they have the capability of appearing to pile up on top of each other in a relaxed, empty bladder, as illustrated in Figure 14.11. As the urinary bladder fills, the epithelial layer unfolds and expands to hold the volume of urine introduced into it. As the bladder fills, it expands and the lining becomes thinner. In other words, the tissue transitions from thick to thin.

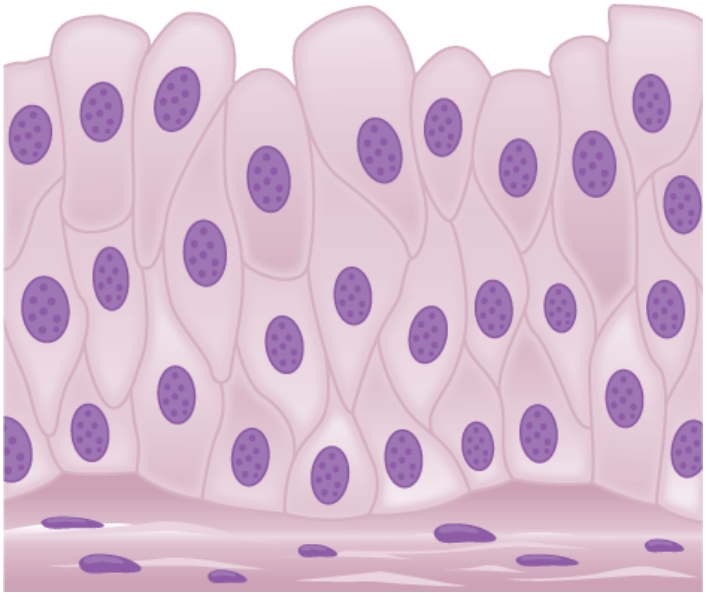


Figure 14.11. Transitional epithelia of the urinary bladder undergo changes in thickness depending on how full the bladder is.

Which of the following statements about types of epithelial cells is false?

1. Simple columnar epithelial cells line the tissue of the lung.
2. Simple cuboidal epithelial cells are involved in the filtering of blood in the kidney.
3. Pseudostratisfied columnar epithilia occur in a single layer, but the arrangement of nuclei makes it appear that more than one layer is present.
4. Transitional epithelia change in thickness depending on how full the bladder is.

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 of the tissues. Some tissues have specialized cells that are not found in the others. The **matrix** in connective tissues gives the tissue its density. When a connective tissue has a high concentration of cells or fibers, it has proportionally a less dense matrix.

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 14.3.

Table 14.3. 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
adipose	adipocytes	few	adipose (fat)
blood	red blood cells, white blood cells	none	blood

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 14.12, loose connective tissue has some fibroblasts; macrophages are present as well. 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, and comprises membranes.

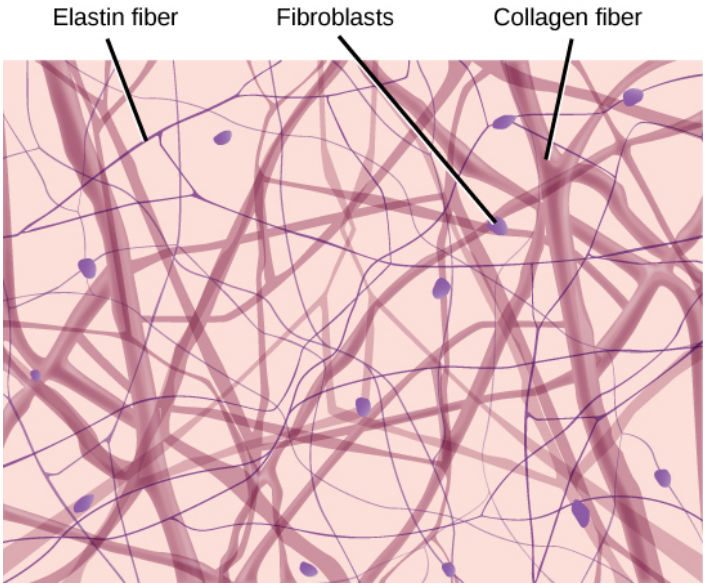


Figure 14.12. Loose connective tissue is composed of loosely woven collagen and elastic fibers. The fibers and other components of the connective tissue matrix are secreted by fibroblasts.

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 14.13, is found in tendons (which connect muscles to bones) and ligaments (which connect bones to bones).

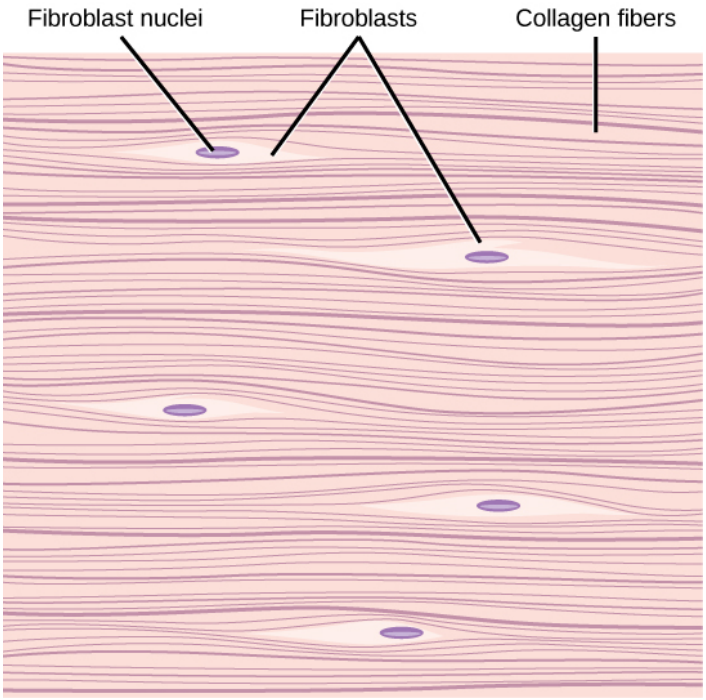


Figure 14.13. Fibrous connective tissue from the tendon has strands of collagen fibers lined up in parallel.

Cartilage

Cartilage is a connective tissue with a large amount of the 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 called **lacunae**.

A cartilage with few collagen and elastic fibers is hyaline cartilage, illustrated in Figure 14.14. 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.

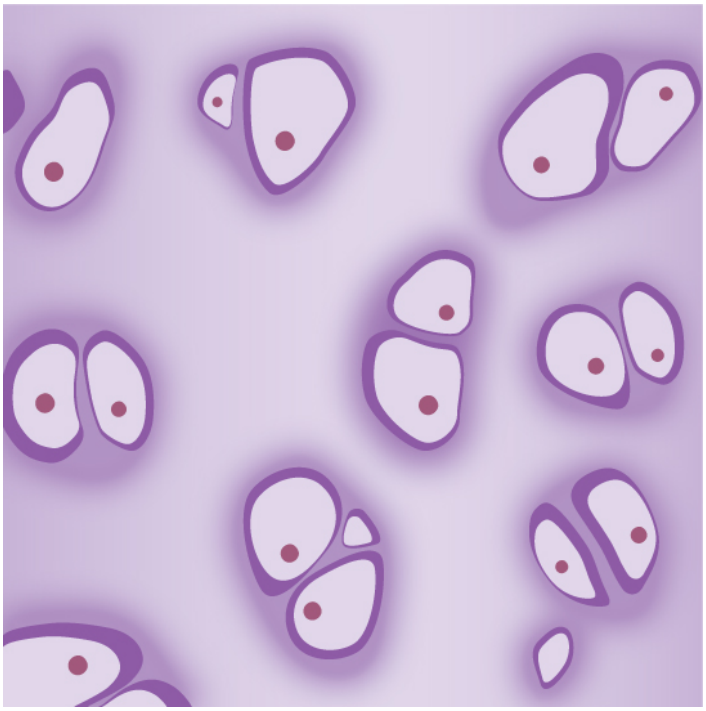


Figure 14.14. Hyaline cartilage consists of a matrix with cells called chondrocytes embedded in it. The chondrocytes exist in cavities in the matrix called lacunae.

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 salts—that give the tissue hardness. Without adequate organic material in the matrix, the tissue breaks; without adequate inorganic material in the matrix, the tissue bends.

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 14.15. 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. Bone tissue forms the internal skeleton of vertebrate animals, providing structure to the animal and points of attachment for tendons.

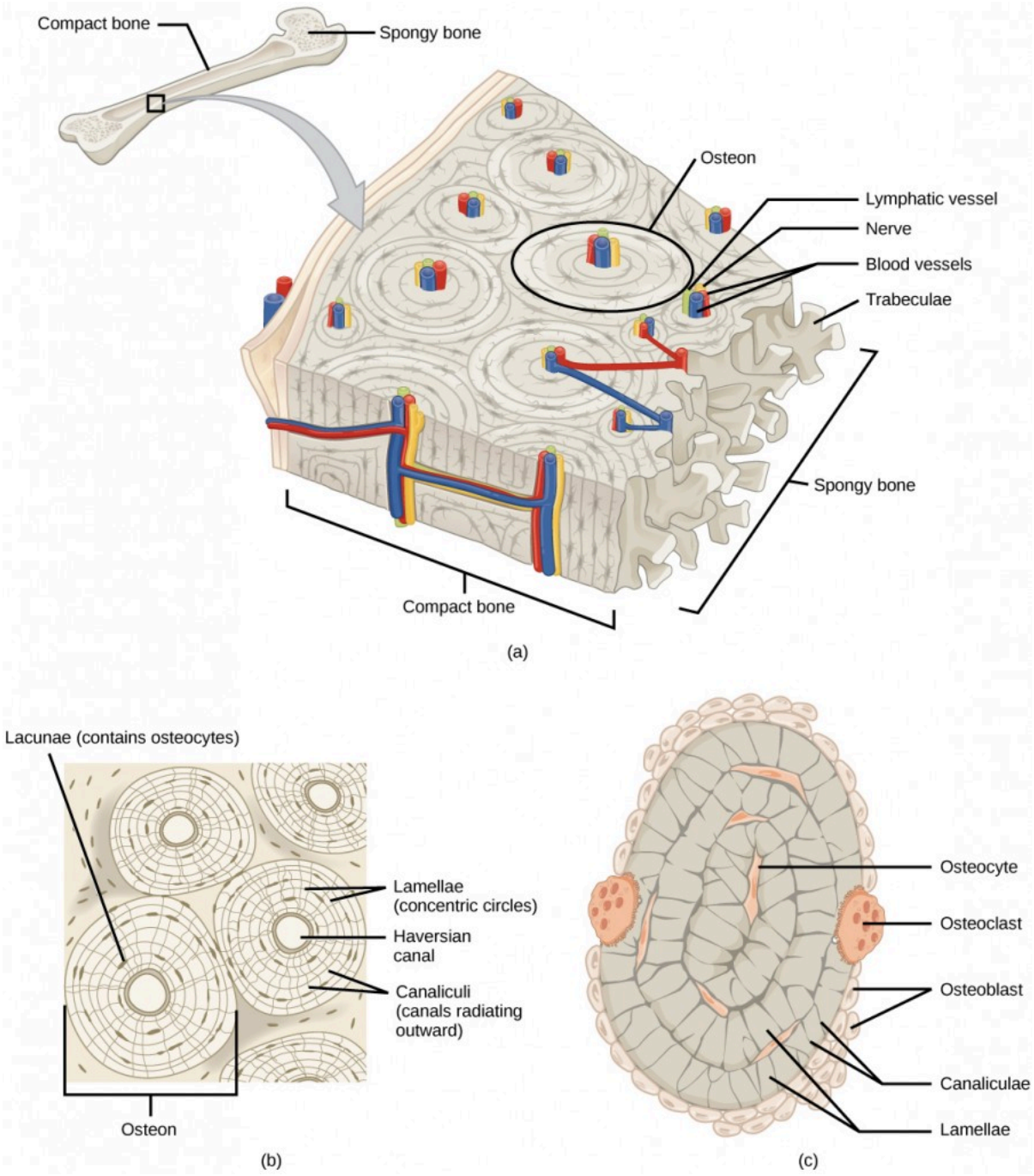


Figure 14.15. (a) Compact bone is a dense matrix on the outer surface of bone. Spongy bone, inside the compact bone, is porous with web-like trabeculae. (b) Compact bone is organized into rings called osteons. Blood vessels, nerves, and lymphatic vessels are found in the central Haversian canal. Rings 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 that

collect and store fat in the form of triglycerides, for energy metabolism. Adipose tissues additionally serve as insulation to help maintain body temperatures, allowing animals to be endothermic, and they function as cushioning against 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 14.16. The thin lines in the image are the cell membranes, and the nuclei are the small, black dots at the edges of the cells.

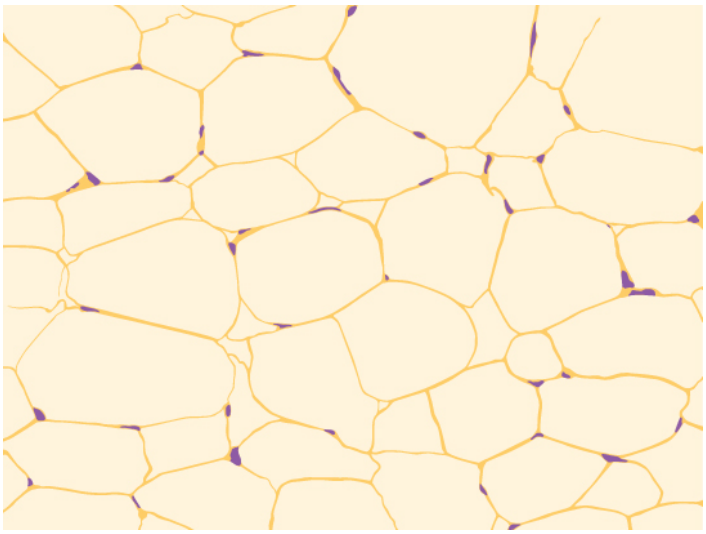


Figure 14.16. Adipose is a connective tissue made up of cells called adipocytes. Adipocytes have small nuclei localized at the cell edge.

Blood

Blood is considered a connective tissue because it has a matrix, as shown in Figure 14.17. The living cell types are red blood cells (RBC), also called erythrocytes, and white blood cells (WBC), also called leukocytes. The fluid portion of whole blood, its matrix, is commonly called plasma.

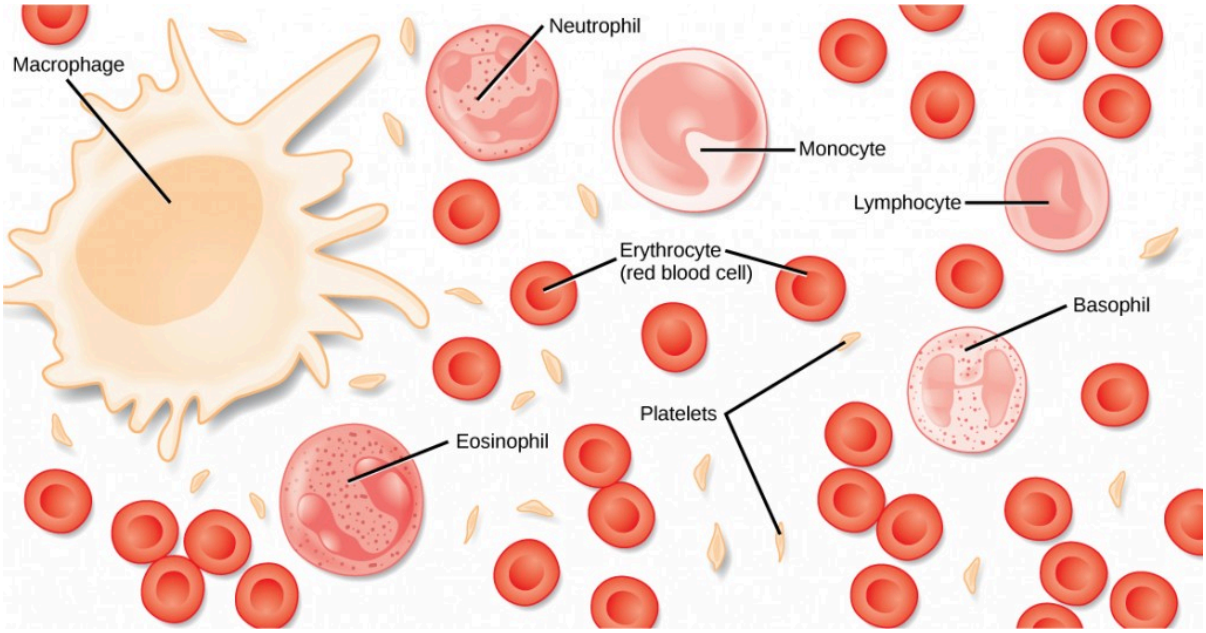


Figure 14.17. Blood is a connective tissue that has a fluid matrix, called plasma, and no fibers. Erythrocytes (red blood cells), the predominant cell type, are involved in the transport of oxygen and carbon dioxide. Also present are various leukocytes (white blood cells) involved in immune response.

The cell found in greatest abundance in blood is the erythrocyte. Erythrocytes are counted in millions in a blood sample: the average number of red blood cells in primates is 4.7 to 5.5 million cells per microliter. Erythrocytes are consistently the same size in a species, but vary in size between species. For example, the average diameter of a primate red blood cell is 7.5 µl, a dog is close at 7.0 µl, but a cat's RBC diameter is 5.9 µl. Sheep erythrocytes are even smaller at 4.6 µl. Mammalian erythrocytes lose their nuclei and mitochondria when they are released from the bone marrow where they are made. Fish, amphibian, and avian red blood cells maintain their nuclei and mitochondria throughout the cell's life. The principal job of an erythrocyte is to carry and deliver oxygen to the tissues.

Leukocytes are the predominant white blood cells found in the peripheral blood. Leukocytes are counted in the thousands in the blood with measurements expressed as ranges: primate counts range from 4,800 to 10,800 cells per µl, dogs from 5,600 to 19,200 cells per µl, cats from 8,000 to 25,000 cells per µl, cattle from 4,000 to 12,000 cells per µl, and pigs from 11,000 to 22,000 cells per µl.

Lymphocytes function primarily in the immune response to foreign antigens or material. Different types of lymphocytes make antibodies tailored to the foreign antigens and control the production of those antibodies. Neutrophils are phagocytic cells and they participate in one of the early lines of defense against microbial invaders, aiding in the removal of bacteria that has entered the body. 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 is a cytoplasmic fragment of a cell in the bone marrow. This is called a platelet or thrombocyte. Platelets participate in the stages leading up to coagulation of the blood to stop bleeding through damaged blood vessels. Blood has a number of functions, but

primarily it transports material through the body to bring nutrients to cells and remove waste material from them.

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 14.4 summarizes these differences.

Table 14.4. 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

Smooth Muscle

Smooth muscle does not have striations in its cells. It has a single, centrally located nucleus, as shown in Figure 14.18. 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.

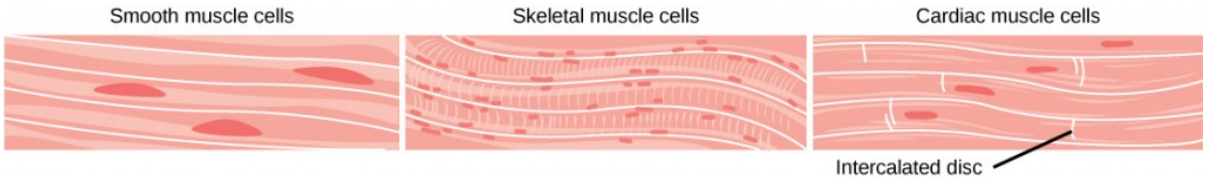


Figure 14.18. Smooth muscle cells do not have striations, while skeletal muscle cells do. Cardiac muscle cells have striations, but, unlike the multinucleate skeletal cells, they have only one nucleus. Cardiac muscle tissue also has intercalated discs, specialized regions running along the plasma membrane 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 14.18 illustrates the histology of skeletal muscle.

Cardiac Muscle

Cardiac muscle, shown in Figure 14.18, 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. An added feature to cardiac muscle cells is a line than extends along the end of the cell as it abuts the next cardiac cell in the row. This line is called an intercalated disc: it assists in passing electrical impulse efficiently from one cell to the next and maintains the strong connection between neighboring cardiac cells.

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 14.19. 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 consists of neurons and glial cells.

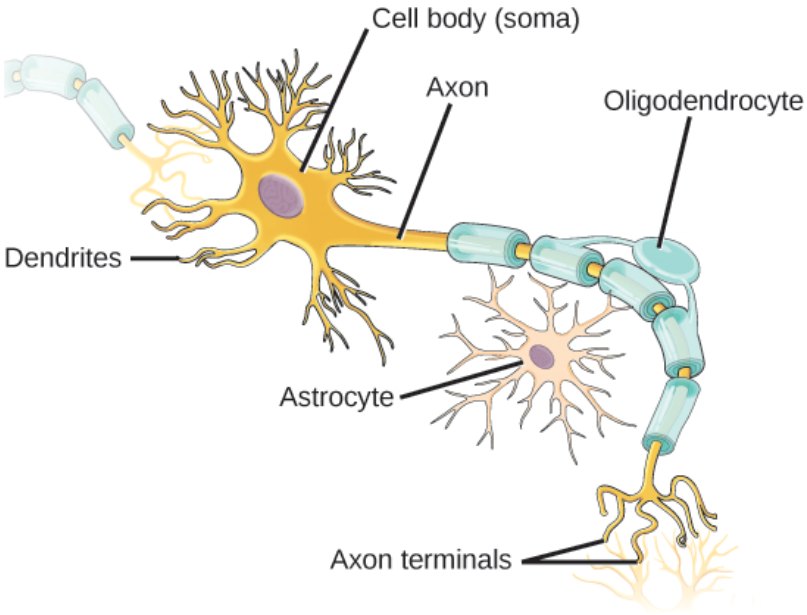


Figure 14.19. The neuron has projections called dendrites that receive signals and projections called axons that send signals. Also shown are two types of glial cells: astrocytes regulate the chemical environment of the nerve cell, and oligodendrocytes insulate the axon so the electrical nerve impulse is transferred more efficiently.

Concept in Action

Click through the interactive review to learn more about epithelial tissues.

Pathologist

A pathologist is a medical doctor or veterinarian who has specialized in the laboratory detection of disease in animals, including humans. These professionals complete medical school education and follow it with an extensive post-graduate residency at a medical center. A pathologist may oversee clinical laboratories for the evaluation of body tissue and blood samples for the detection of disease or infection. They examine tissue specimens through a microscope to identify cancers and other diseases. Some pathologists perform autopsies to determine the cause of death and the progression of disease.

Exercises

1. Which of the following statements about types of epithelial cells is false?
    1. Simple columnar epithelial cells line the tissue of the lung.
    2. Simple cuboidal epithelial cells are involved in the filtering of blood in the kidney.
    3. Pseudostratisfied columnar epithilia occur in a single layer, but the arrangement of nuclei makes it appear that more than one layer is present.
    4. Transitional epithelia change in thickness depending on how full the bladder is.
2. State whether each of the following processes are regulated by a positive feedback loop or a negative feedback loop.
    1. A person feels satiated after eating a large meal.
    2. The blood has plenty of red blood cells. As a result, erythropoietin, a hormone that stimulates the production of new red blood cells, is no longer released from the kidney.
3. When bacteria are destroyed by leuckocytes, pyrogens are released into the blood. Pyrogens reset the body's thermostat to a higher temperature, resulting in fever. How might pyrogens cause the body temperature to rise?
4. Which type of animal maintains a constant internal body temperature?
    1. endotherm
    2. ectotherm
    3. coelomate
    4. mesoderm
5. The symmetry found in animals that move swiftly is _____.
    1. radial
    2. bilateral
    3. sequential
    4. interrupted
6. What term describes the condition of a desert mouse that lowers its metabolic rate and "sleeps" during the hot day?
    1. turgid
    2. hibernation
    3. estivation
    4. normal sleep pattern
7. A plane that divides an animal into equal right and left portions is _____.
    1. diagonal
    2. midsagittal

3. coronal
4. transverse

8. A plane that divides an animal into dorsal and ventral portions is _____.

    1. sagittal
    2. midsagittal
    3. coronal
    4. transverse

9. The pleural cavity is a part of which cavity?

    1. dorsal cavity
    2. thoracic cavity
    3. abdominal cavity
    4. pericardial cavity

10. Which type of epithelial cell is best adapted to aid diffusion?

    1. squamous
    2. cuboidal
    3. columnar
    4. transitional

11. Which type of epithelial cell is found in glands?

    1. squamous
    2. cuboidal
    3. columnar
    4. transitional

12. Which type of epithelial cell is found in the urinary bladder?

    1. squamous
    2. cuboidal
    3. columnar
    4. transitional

13. Which type of connective tissue has the most fibers?

    1. loose connective tissue
    2. fibrous connective tissue
    3. cartilage
    4. bone

14. Which type of connective tissue has a mineralized different matrix?
    1. loose connective tissue
    2. fibrous connective tissue
    3. cartilage
    4. bone
15. The cell found in bone that breaks it down is called an _____.
    1. osteoblast
    2. osteocyte
    3. osteoclast
    4. osteon
16. The cell found in bone that makes the bone is called an _____.
    1. osteoblast
    2. osteocyte
    3. osteoclast
    4. osteon
17. Plasma is the _____.
    1. fibers in blood
    2. matrix of blood
    3. cell that phagocytizes bacteria
    4. cell fragment found in the tissue
18. The type of muscle cell under voluntary control is the _____.
    1. smooth muscle
    2. skeletal muscle
    3. cardiac muscle
    4. visceral muscle
19. The part of a neuron that contains the nucleus is the
    1. cell body
    2. dendrite
    3. axon
    4. glial
20. When faced with a sudden drop in environmental temperature, an endothermic animal will:
    1. experience a drop in its body temperature

2. wait to see if it goes lower
3. increase muscle activity to generate heat
4. add fur or fat to increase insulation

21. Which is an example of negative feedback?

    1. lowering of blood glucose after a meal
    2. blood clotting after an injury
    3. lactation during nursing
    4. uterine contractions during labor

22. Which method of heat exchange occurs during direct contact between the source and animal?

    1. radiation
    2. evaporation
    3. convection
    4. conduction

23. The body's thermostat is located in the _____.

    1. homeostatic receptor
    2. hypothalamus
    3. medulla
    4. vasodilation center

24. How does diffusion limit the size of an organism? How is this counteracted?
25. What is the relationship between BMR and body size? Why?
26. How can squamous epithelia both facilitate diffusion and prevent damage from abrasion?
27. What are the similarities between cartilage and bone?
28. Why are negative feedback loops used to control body homeostasis?
29. Why is a fever a "good thing" during a bacterial infection?
30. How is a condition such as diabetes a good example of the failure of a set point in humans?

Answers

1. A
2. Both processes are the result of negative feedback loops. Negative feedback loops, which tend to keep a system at equilibrium, are more common than positive feedback loops.
3. Pyrogens increase body temperature by causing the blood vessels to constrict, inducing shivering, and stopping sweat glands from secreting fluid.
4. A
5. B

6. C
7. B
8. D
9. B
10. C
11. B
12. D
13. B
14. D
15. C
16. A
17. B
18. B
19. B
20. C
21. A
22. D
23. B
24. Diffusion is effective over a very short distance. If a cell exceeds this distance in its size, the center of the cell cannot get adequate nutrients nor can it expel enough waste to survive. To compensate for this, cells can loosely adhere to each other in a liquid medium, or develop into multi-celled organisms that use circulatory and respiratory systems to deliver nutrients and remove wastes.
25. Basal Metabolic Rate is an expression of the metabolic processes that occur to maintain an individual's functioning and body temperature. Smaller bodied animals have a relatively large surface area compared to a much larger animal. The large animal's large surface area leads to increased heat loss that the animal must compensate for, resulting in a higher BMR. A small animal, having less relative surface area, does not lose as much heat and has a correspondingly lower BMR.
26. Squamous epithelia can be either simple or stratified. As a single layer of cells, it presents a very thin epithelia that minimally inhibits diffusion. As a stratified epithelia, the surface cells can be sloughed off and the cells in deeper layers protect the underlying tissues from damage.
27. Both contain cells other than the traditional fibroblast. Both have cells that lodge in spaces within the tissue called lacunae. Both collagen and elastic fibers are found in bone and cartilage. Both tissues participate in vertebrate skeletal development and formation.
28. An adjustment to a change in the internal or external environment requires a change in the direction of the stimulus. A negative feedback loop accomplishes this, while a positive feedback loop would continue the stimulus and result in harm to the animal.
29. Mammalian enzymes increase activity to the point of denaturation, increasing the chemical

activity of the cells involved. Bacterial enzymes have a specific temperature for their most efficient activity and are inhibited at either higher or lower temperatures. Fever results in an increase in the destruction of the invading bacteria by increasing the effectiveness of body defenses and an inhibiting bacterial metabolism.

30. Diabetes is often associated with a lack in production of insulin. Without insulin, blood glucose levels go up after a meal, but never go back down to normal levels.

Glossary

cartilage: type of connective tissue with a large amount of ground substance matrix, cells called chondrocytes, and some amount of fibers
chondrocyte: cell found in cartilage
columnar epithelia: epithelia made of cells taller than they are wide, specialized in absorption
connective tissue: type of tissue made of cells, ground substance matrix, and fibers
cuboidal epithelia: epithelia made of cube-shaped cells, specialized in glandular functions
dorsal cavity: body cavity on the posterior or back portion of an animal; includes the cranial and vertebral cavities
ectotherm: animal incapable of maintaining a relatively constant internal body temperature
endotherm: animal capable of maintaining a relatively constant internal body temperature
epithelial tissue: tissue that either lines or covers organs or other tissues
estivation: torpor in response to extremely high temperatures and low water availability
fibrous connective tissue: type of connective tissue with a high concentration of fibers
hibernation: torpor over a long period of time, such as a winter
homeostasis: dynamic equilibrium maintaining appropriate body functions
lacuna: space in cartilage and bone that contains living cells
matrix: component of connective tissue made of both living and non-living (ground substances) cells
negative feedback loop: feedback to a control mechanism that increases or decreases a stimulus instead of maintaining it
osteon: subunit of compact bone
positive feedback loop: feedback to a control mechanism that continues the direction of a stimulus
pseudostratified: layer of epithelia that appears multilayered, but is a simple covering
set point: midpoint or target point in homeostasis
simple epithelia: single layer of epithelial cells
squamous epithelia: type of epithelia made of flat cells, specialized in aiding diffusion or preventing abrasion
stratified epithelia: multiple layers of epithelial cells
torpor: decrease in activity and metabolism that allows an animal to survive adverse conditions
trabecula: tiny plate that makes up spongy bone and gives it strength

14.3 Homeostasis

Learning Objectives

By the end of this section, you will be able to:

- Define homeostasis
- Describe the factors affecting homeostasis
- Discuss positive and negative feedback mechanisms used in homeostasis
- Describe thermoregulation of endothermic and ectothermic animals

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 deviation parameter toward the set point. For instance, if the body becomes too warm, adjustments are made to cool the animal. If the blood’s glucose rises after a meal, adjustments are made to lower the blood glucose level by getting the nutrient into tissues that need it or to store it for later use.

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.

Negative Feedback Mechanisms

Any homeostatic process that changes the direction of the stimulus is a **negative feedback loop**. It may either increase or decrease the stimulus, but the stimulus is not allowed to continue as it did before the receptor sensed it. 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. 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 14.20. 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, but not in the direction expected by the use of the term “negative.” Another example of an increase as a result of the feedback loop is the control of blood calcium. If calcium levels decrease, specialized cells in the parathyroid gland sense this and release parathyroid hormone (PTH), causing an increased absorption of calcium through the intestines and kidneys and, possibly, the breakdown of bone in order to liberate calcium. The effects of PTH are to raise blood levels of the element. Negative feedback loops are the predominant mechanism used in homeostasis.

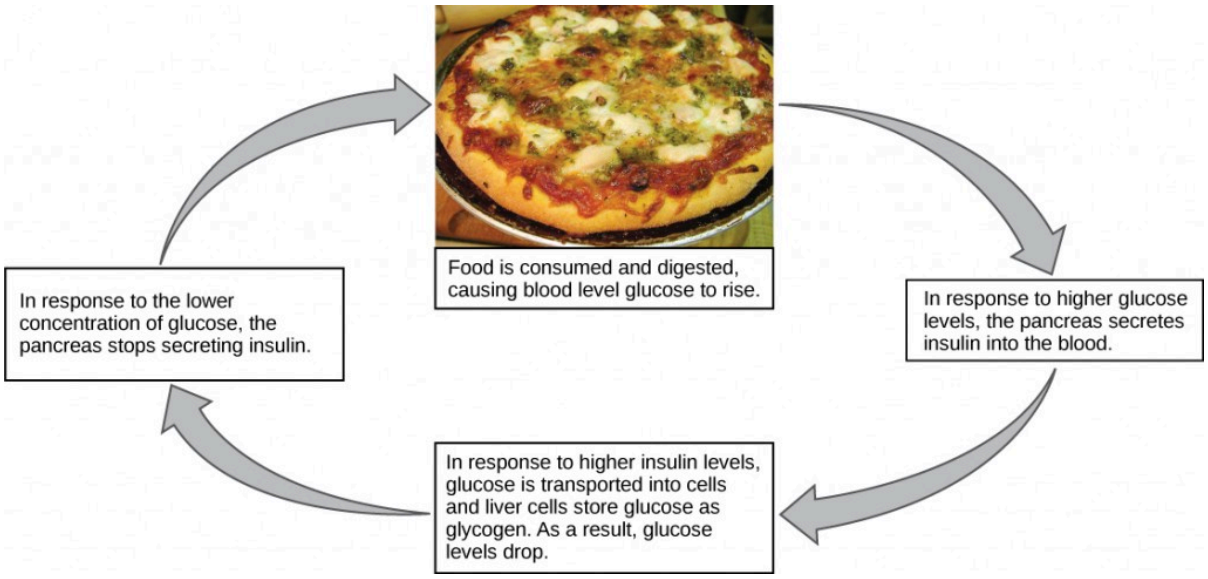


Figure 14.20. Blood sugar levels are controlled by a negative feedback loop. (credit: modification of work by Jon Sullivan)

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 14.21. The hormone oxytocin, made by the endocrine system, stimulates the contraction of the uterus. This produces pain sensed by the nervous system. Instead of lowering the

oxytocin and causing the pain to subside, more oxytocin is produced until the contractions are powerful enough to produce childbirth.

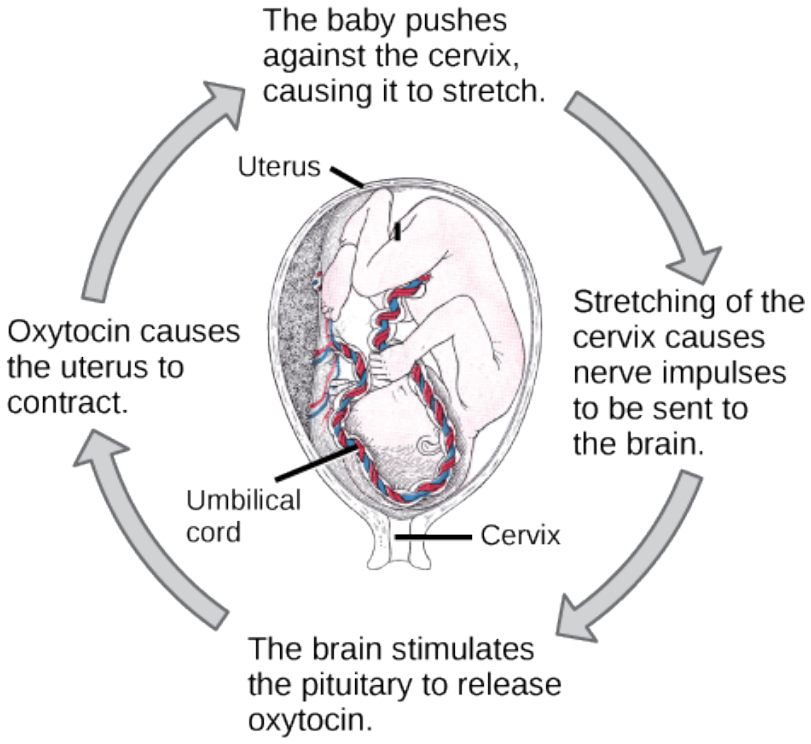


Figure 14.21. The birth of a human infant is the result of positive feedback.

State whether each of the following processes is regulated by a positive feedback loop or a negative feedback loop.

1. A person feels satiated after eating a large meal.
2. The blood has plenty of red blood cells. As a result, erythropoietin, a hormone that stimulates the production of new red blood cells, is no longer released from the kidney.

Set Point

It is possible to adjust a system's set point. 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. This is called a process of **alteration** of the set point in a feedback loop.

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.

Concept in Action

Feedback mechanisms can be understood in terms of driving a race car along a track: watch a short video lesson on positive and negative feedback loops.

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 with high heat (around 50^{o}C for mammals). Enzyme activity will 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.

Concept in Action

Watch this Discovery Channel video on thermoregulation to see illustrations of this process in a variety of animals.

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 a homeotherm. 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, which an ectotherm cannot due to differing enzyme levels of activity.

Heat can be exchanged between an animal and its environment through four mechanisms: radiation, evaporation, convection, and conduction (Figure 14.22). 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.

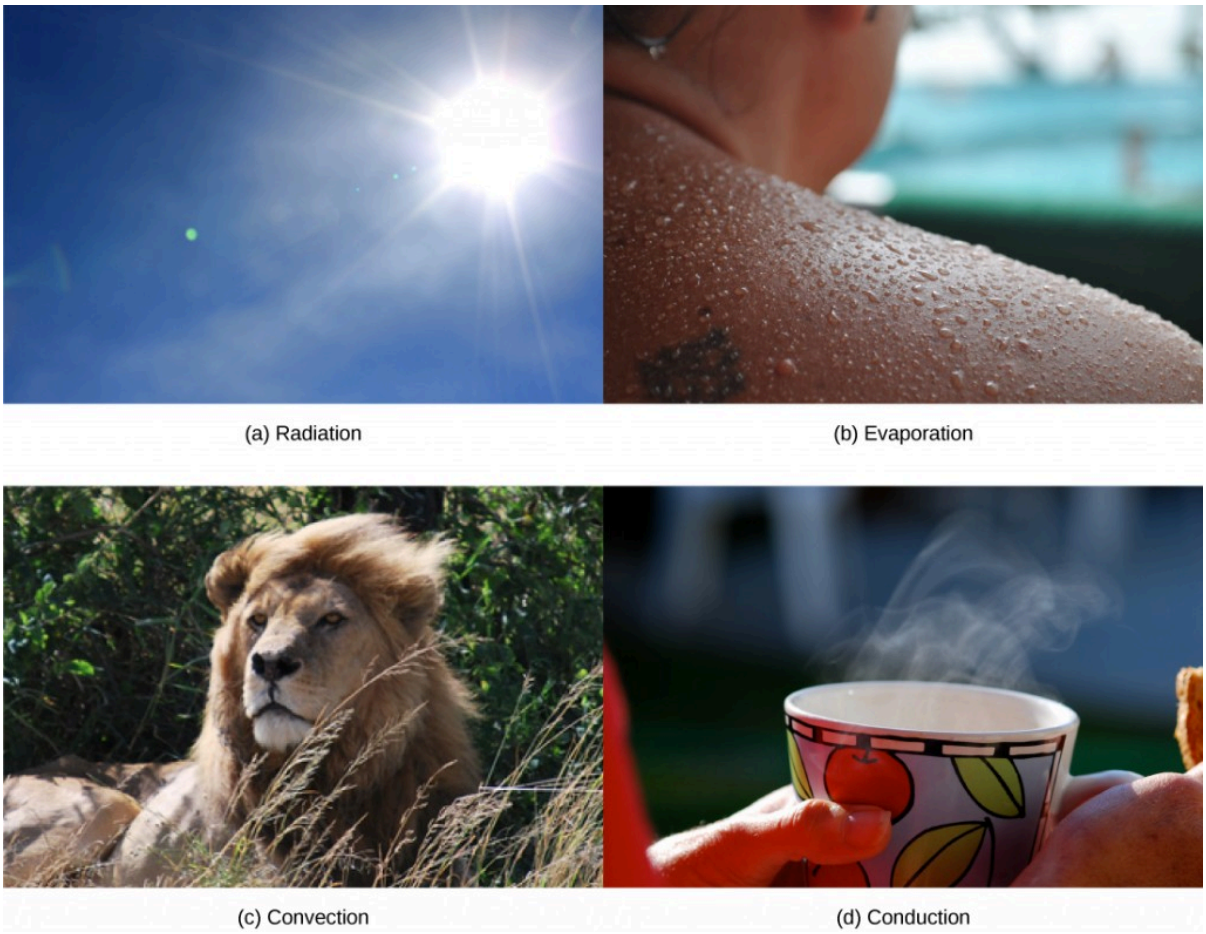


Figure 14.22. Heat can be exchanged by four mechanisms: (a) radiation, (b) evaporation, (c) convection, or (d) conduction. (credit b: modification of work by "Kullez"/Flickr; credit c: modification of work by Chad Rosenthal; credit d: modification of work by "stacey.d"/Flickr)

Heat Conservation and Dissipation

Animals conserve or dissipate heat in a variety of ways. In certain climates, endothermic animals have some form of insulation, such as fur, fat, feathers, or some combination thereof. Animals with thick fur or feathers create an insulating layer of air between their skin and internal organs. Polar bears and seals

live and swim in a subfreezing environment and yet maintain a constant, warm, body temperature. The arctic fox, for example, uses its fluffy tail as extra insulation when it curls up to sleep in cold weather. Mammals have a residual effect from shivering and increased muscle activity: arrector pili muscles cause “goose bumps,” causing small hairs to stand up when the individual is cold; this has the intended effect of increasing body temperature. Mammals use layers of fat to achieve the same end. Loss of significant amounts of body fat will compromise an individual’s ability to conserve heat.

Endotherms use their circulatory systems to help maintain body temperature. Vasodilation brings more blood and heat to the body surface, facilitating radiation and evaporative heat loss, which helps to cool the body. Vasoconstriction reduces blood flow in peripheral blood vessels, forcing blood toward the core and the vital organs found there, and conserving heat. Some animals have adaptions to their circulatory system that enable them to transfer heat from arteries to veins, warming blood returning to the heart. This is called a countercurrent heat exchange; it prevents the cold venous blood from cooling the heart and other internal organs. This adaption can be shut down in some animals to prevent overheating the internal organs. The countercurrent adaption is found in many animals, including dolphins, sharks, bony fish, bees, and hummingbirds. In contrast, similar adaptations can help cool endotherms when needed, such as dolphin flukes and elephant ears.

Some ectothermic animals use changes in their behavior to help regulate body temperature. For example, a desert ectothermic animal may simply seek cooler areas during the hottest part of the day in the desert to keep from getting too warm. The same animals may climb onto rocks to capture heat during a cold desert night. Some animals seek water to aid evaporation in cooling them, as seen with reptiles. Other ectotherms use group activity such as the activity of bees to warm a hive to survive winter.

Many animals, especially mammals, use metabolic waste heat as a heat source. When muscles are contracted, most of the energy from the ATP used in muscle actions is wasted energy that translates into heat. Severe cold elicits a shivering reflex that generates heat for the body. Many species also have a type of adipose tissue called brown fat that specializes in generating heat.

Neural Control of Thermoregulation

The nervous system is important to **thermoregulation**, as illustrated in Figure 14.23. The processes of homeostasis and temperature control are centered in the hypothalamus of the advanced animal brain.

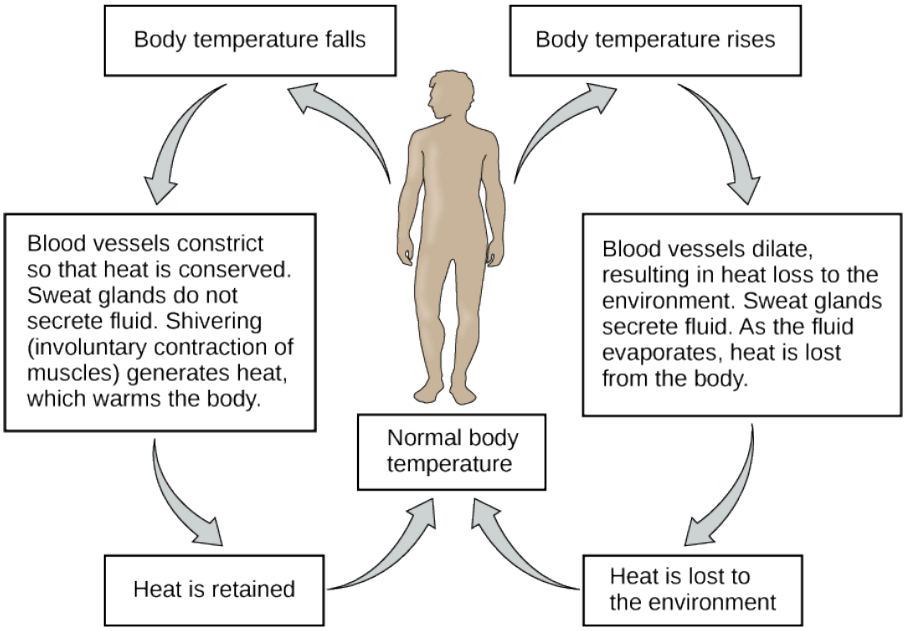


Figure 14.23. The body is able to regulate temperature in response to signals from the nervous system.

When bacteria are destroyed by leuckocytes, pyrogens are released into the blood. Pyrogens reset the body's thermostat to a higher temperature, resulting in fever. How might pyrogens cause the body temperature to rise?

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 are released into the blood. These pyrogens 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, which reduces a nutrient needed by 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. A fever that was once thought to be a complication of an infection is now understood to be a normal defense mechanism.

Exercises

1. Which of the following statements about types of epithelial cells is false?
    1. Simple columnar epithelial cells line the tissue of the lung.
    2. Simple cuboidal epithelial cells are involved in the filtering of blood in the kidney.
    3. Pseudostratisfied columnar epithilia occur in a single layer, but the arrangement of nuclei makes it appear that more than one layer is present.
    4. Transitional epithelia change in thickness depending on how full the bladder is.
2. State whether each of the following processes are regulated by a positive feedback loop or a negative feedback loop.
    1. A person feels satiated after eating a large meal.
    2. The blood has plenty of red blood cells. As a result, erythropoietin, a hormone that stimulates the production of new red blood cells, is no longer released from the kidney.
3. When bacteria are destroyed by leuckocytes, pyrogens are released into the blood. Pyrogens reset the body's thermostat to a higher temperature, resulting in fever. How might pyrogens cause the body temperature to rise?
4. Which type of animal maintains a constant internal body temperature?
    1. endotherm
    2. ectotherm
    3. coelomate
    4. mesoderm
5. The symmetry found in animals that move swiftly is _____.
    1. radial
    2. bilateral
    3. sequential
    4. interrupted
6. What term describes the condition of a desert mouse that lowers its metabolic rate and "sleeps" during the hot day?
    1. turgid
    2. hibernation
    3. estivation
    4. normal sleep pattern
7. A plane that divides an animal into equal right and left portions is _____.
    1. diagonal
    2. midsagittal

3. coronal
4. transverse

8. A plane that divides an animal into dorsal and ventral portions is _____.
    1. sagittal
    2. midsagittal
    3. coronal
    4. transverse
9. The pleural cavity is a part of which cavity?
    1. dorsal cavity
    2. thoracic cavity
    3. abdominal cavity
    4. pericardial cavity
10. Which type of epithelial cell is best adapted to aid diffusion?
    1. squamous
    2. cuboidal
    3. columnar
    4. transitional
11. Which type of epithelial cell is found in glands?
    1. squamous
    2. cuboidal
    3. columnar
    4. transitional
12. Which type of epithelial cell is found in the urinary bladder?
    1. squamous
    2. cuboidal
    3. columnar
    4. transitional
13. Which type of connective tissue has the most fibers?
    1. loose connective tissue
    2. fibrous connective tissue
    3. cartilage
    4. bone

14. Which type of connective tissue has a mineralized different matrix?
    1. loose connective tissue
    2. fibrous connective tissue
    3. cartilage
    4. bone
15. The cell found in bone that breaks it down is called an _____.
    1. osteoblast
    2. osteocyte
    3. osteoclast
    4. osteon
16. The cell found in bone that makes the bone is called an _____.
    1. osteoblast
    2. osteocyte
    3. osteoclast
    4. osteon
17. Plasma is the _____.
    1. fibers in blood
    2. matrix of blood
    3. cell that phagocytizes bacteria
    4. cell fragment found in the tissue
18. The type of muscle cell under voluntary control is the _____.
    1. smooth muscle
    2. skeletal muscle
    3. cardiac muscle
    4. visceral muscle
19. The part of a neuron that contains the nucleus is the
    1. cell body
    2. dendrite
    3. axon
    4. glial
20. When faced with a sudden drop in environmental temperature, an endothermic animal will:
    1. experience a drop in its body temperature

2. wait to see if it goes lower
3. increase muscle activity to generate heat
4. add fur or fat to increase insulation

21. Which is an example of negative feedback?

    1. lowering of blood glucose after a meal
    2. blood clotting after an injury
    3. lactation during nursing
    4. uterine contractions during labor

22. Which method of heat exchange occurs during direct contact between the source and animal?

    1. radiation
    2. evaporation
    3. convection
    4. conduction

23. The body's thermostat is located in the _____.

    1. homeostatic receptor
    2. hypothalamus
    3. medulla
    4. vasodilation center

24. How does diffusion limit the size of an organism? How is this counteracted?
25. What is the relationship between BMR and body size? Why?
26. How can squamous epithelia both facilitate diffusion and prevent damage from abrasion?
27. What are the similarities between cartilage and bone?
28. Why are negative feedback loops used to control body homeostasis?
29. Why is a fever a "good thing" during a bacterial infection?
30. How is a condition such as diabetes a good example of the failure of a set point in humans?

Answers

1. A
2. Both processes are the result of negative feedback loops. Negative feedback loops, which tend to keep a system at equilibrium, are more common than positive feedback loops.
3. Pyrogens increase body temperature by causing the blood vessels to constrict, inducing shivering, and stopping sweat glands from secreting fluid.
4. A
5. B

6. C
7. B
8. D
9. B
10. C
11. B
12. D
13. B
14. D
15. C
16. A
17. B
18. B
19. B
20. C
21. A
22. D
23. B
24. Diffusion is effective over a very short distance. If a cell exceeds this distance in its size, the center of the cell cannot get adequate nutrients nor can it expel enough waste to survive. To compensate for this, cells can loosely adhere to each other in a liquid medium, or develop into multi-celled organisms that use circulatory and respiratory systems to deliver nutrients and remove wastes.
25. Basal Metabolic Rate is an expression of the metabolic processes that occur to maintain an individual's functioning and body temperature. Smaller bodied animals have a relatively large surface area compared to a much larger animal. The large animal's large surface area leads to increased heat loss that the animal must compensate for, resulting in a higher BMR. A small animal, having less relative surface area, does not lose as much heat and has a correspondingly lower BMR.
26. Squamous epithelia can be either simple or stratified. As a single layer of cells, it presents a very thin epithelia that minimally inhibits diffusion. As a stratified epithelia, the surface cells can be sloughed off and the cells in deeper layers protect the underlying tissues from damage.
27. Both contain cells other than the traditional fibroblast. Both have cells that lodge in spaces within the tissue called lacunae. Both collagen and elastic fibers are found in bone and cartilage. Both tissues participate in vertebrate skeletal development and formation.
28. An adjustment to a change in the internal or external environment requires a change in the direction of the stimulus. A negative feedback loop accomplishes this, while a positive feedback loop would continue the stimulus and result in harm to the animal.
29. Mammalian enzymes increase activity to the point of denaturation, increasing the chemical

activity of the cells involved. Bacterial enzymes have a specific temperature for their most efficient activity and are inhibited at either higher or lower temperatures. Fever results in an increase in the destruction of the invading bacteria by increasing the effectiveness of body defenses and an inhibiting bacterial metabolism.

30. Diabetes is often associated with a lack in production of insulin. Without insulin, blood glucose levels go up after a meal, but never go back down to normal levels.

Glossary

acclimatization: alteration in a body system in response to environmental change
alteration: change of the set point in a homeostatic system
cartilage: type of connective tissue with a large amount of ground substance matrix, cells called chondrocytes, and some amount of fibers
chondrocyte: cell found in cartilage
columnar epithelia: epithelia made of cells taller than they are wide, specialized in absorption
connective tissue: type of tissue made of cells, ground substance matrix, and fibers
cuboidal epithelia: epithelia made of cube-shaped cells, specialized in glandular functions
dorsal cavity: body cavity on the posterior or back portion of an animal; includes the cranial and vertebral cavities
ectotherm: animal incapable of maintaining a relatively constant internal body temperature
endotherm: animal capable of maintaining a relatively constant internal body temperature
estivation: torpor in response to extremely high temperatures and low water availability
fibrous connective tissue: type of connective tissue with a high concentration of fibers
hibernation: torpor over a long period of time, such as a winter
homeostasis: dynamic equilibrium maintaining appropriate body functions
lacuna: space in cartilage and bone that contains living cells
matrix: component of connective tissue made of both living and non-living (ground substances) cells
negative feedback loop: feedback to a control mechanism that increases or decreases a stimulus instead of maintaining it
osteon: subunit of compact bone
positive feedback loop: feedback to a control mechanism that continues the direction of a stimulus
set point: midpoint or target point in homeostasis
squamous epithelia: type of epithelia made of flat cells, specialized in aiding diffusion or preventing abrasion
stratified epithelia: multiple layers of epithelial cells
thermoregulation: regulation of body temperature
torpor: decrease in activity and metabolism that allows an animal to survive adverse conditions

Chapter 15. Animal Nutrition and the Digestive System

Figure 15.1. For humans, fruits and vegetables are important in maintaining a balanced diet. (credit: modification of work by Julie Rybarczyk)

Introduction

All living organisms need nutrients to survive. While plants can obtain the molecules required for cellular function through the process of photosynthesis, most animals obtain their nutrients by the consumption of other organisms. At the cellular level, the biological molecules necessary for animal function are amino acids, lipid molecules, nucleotides, and simple sugars. However, the food consumed consists of protein, fat, and complex carbohydrates. Animals must convert these macromolecules into the simple molecules required for maintaining cellular functions, such as assembling new molecules, cells, and tissues. The conversion of the food consumed to the nutrients required is a multi-step process involving digestion and absorption. During digestion, food particles are broken down to smaller components, and later, they are absorbed by the body.

One of the challenges in human nutrition is maintaining a balance between food intake, storage, and energy expenditure. Imbalances can have serious health consequences. For example, eating too much food while not expending much energy leads to obesity, which in turn will increase the risk of developing illnesses such as type-2 diabetes and cardiovascular disease. The recent rise in obesity and related diseases makes understanding the role of diet and nutrition in maintaining good health all the more important.

15.1 Digestive Systems

Learning Objectives

By the end of this section, you will be able to:

- Explain the processes of digestion and absorption
- Compare and contrast different types of digestive systems
- Explain the specialized functions of the organs involved in processing food in the body
- Describe the ways in which organs work together to digest food and absorb nutrients

Animals 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 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 15.2 include vertebrates like deer, koalas, and some bird species, as well as invertebrates such as crickets and caterpillars. These animals have evolved digestive systems capable of handling large amounts of plant material. Herbivores can be further classified into frugivores (fruit-eaters), granivores (seed eaters), nectivores (nectar feeders), and folivores (leaf eaters).

Figure 15.2. Herbivores, like this (a) mule deer and (b) monarch caterpillar, eat primarily plant material. (credit a: modification of work by Bill Ebbesen; credit b: modification of work by Doug Bowman)

Carnivores are animals that eat other animals. The word carnivore is derived from Latin and literally means "meat eater." Wild cats such as lions, shown in Figure 35.3 **a** and tigers are examples of vertebrate carnivores, as are snakes and sharks, while invertebrate carnivores include sea stars, spiders, and ladybugs, shown in Figure 15.3 **b**. 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.

Figure 15.3. Carnivores like the (a) lion eat primarily meat. The (b) ladybug is also a carnivore that consumes small insects called aphids. (credit a: modification of work by Kevin Pluck; credit b: modification of work by Jon Sullivan)

Omnivores are animals that eat both plant- and animal-derived food. In Latin, omnivore means to eat everything. Humans, bears (shown in Figure 15.4 **a**), and chickens are example of vertebrate omnivores; invertebrate omnivores include cockroaches and crayfish (shown in Figure 15.4 **b**).

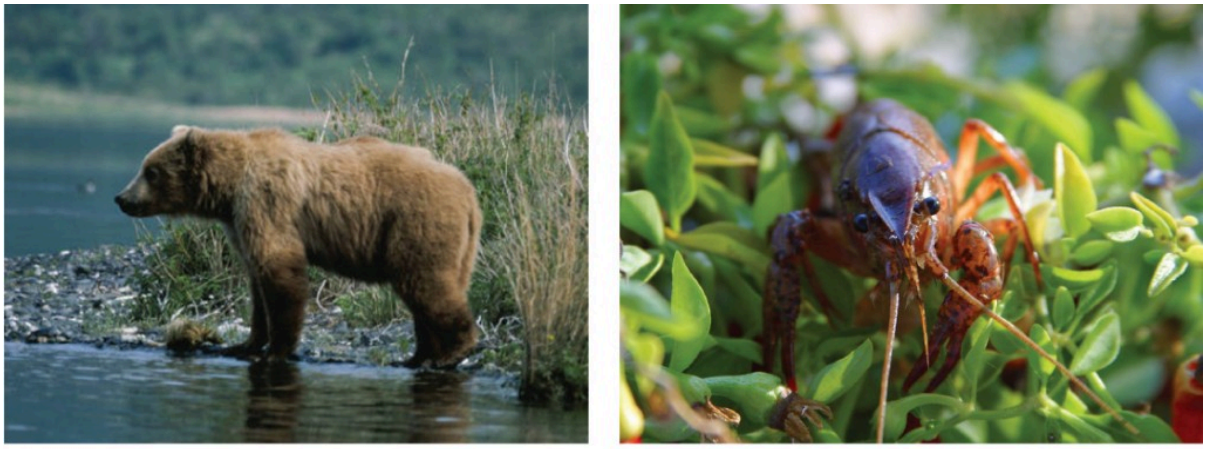


Figure 15.4. Omnivores like the (a) bear and (b) crayfish eat both plant and animal based food. (credit a: modification of work by Dave Menke; credit b: modification of work by Jon Sullivan)

Invertebrate Digestive Systems

Animals have evolved different types of digestive systems to aid in the digestion of the different foods they consume. The simplest example is that of a **gastrovascular cavity** and is found in organisms with only one opening for digestion. Platyhelminthes (flatworms), Ctenophora (comb jellies), and Cnidaria (coral, jelly fish, and sea anemones) use this type of digestion. Gastrovascular cavities, as shown in Figure 15.5 **a**, are typically a blind tube or cavity with only one opening, the “mouth”, which also serves as an “anus”. Ingested material enters the mouth and passes through a hollow, tubular cavity. Cells within the cavity secrete digestive enzymes that break down the food. The food particles are engulfed by the cells lining the gastrovascular cavity.

The **alimentary canal**, shown in Figure 15.5 **b**, is a more advanced system: it consists of one tube with a mouth at one end and an anus at the other. Earthworms are an example of an animal with an alimentary canal. Once the food is ingested through the mouth, it passes through the esophagus and is stored in an organ called the crop; then it passes into the gizzard where it is churned and digested. From the gizzard, the food passes through the intestine, the nutrients are absorbed, and the waste is eliminated as feces, called castings, through the anus.

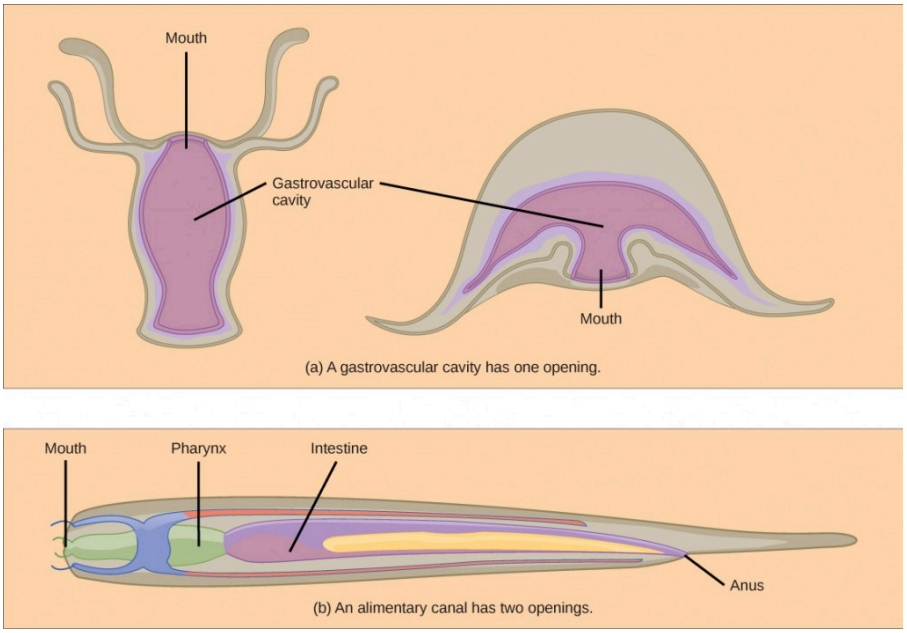


Figure 15.5. (a) A gastrovascular cavity has a single opening through which food is ingested and waste is excreted, as shown in this hydra and in this jellyfish medusa. (b) An alimentary canal has two openings: a mouth for ingesting food, and an anus for eliminating waste, as shown in this nematode.

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 15.6 **ab**. 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.

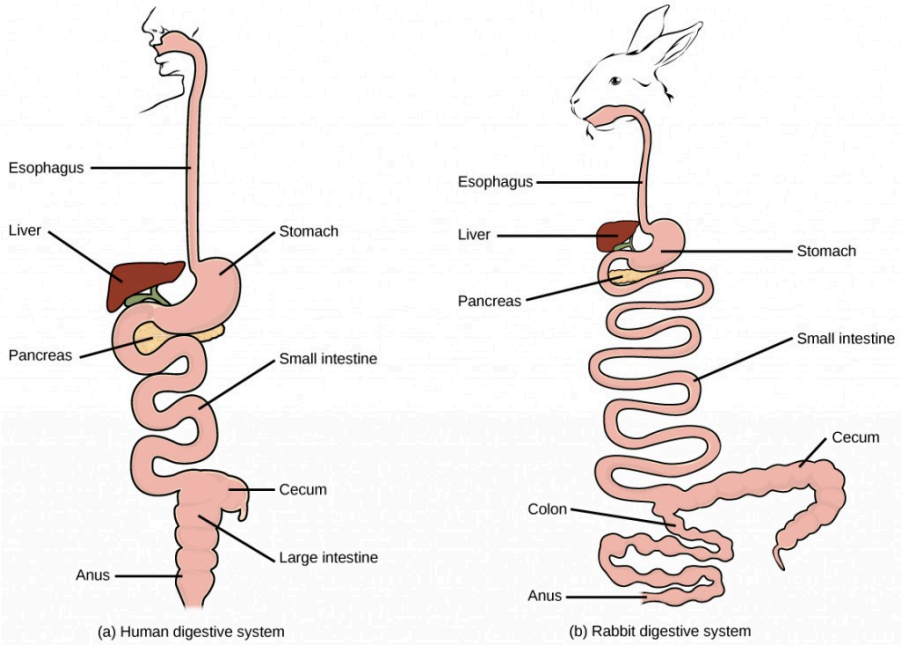


Figure 15.6.
(a) Humans and herbivores, such as the (b) rabbit, have a monogastric digestive system. However, in the rabbit the small intestine and cecum are enlarged to allow more time to digest plant material. The enlarged organ provides more surface area for absorption of nutrients. Rabbits digest 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. They do not have teeth and so their digestive system, shown in Figure 15.7, 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.

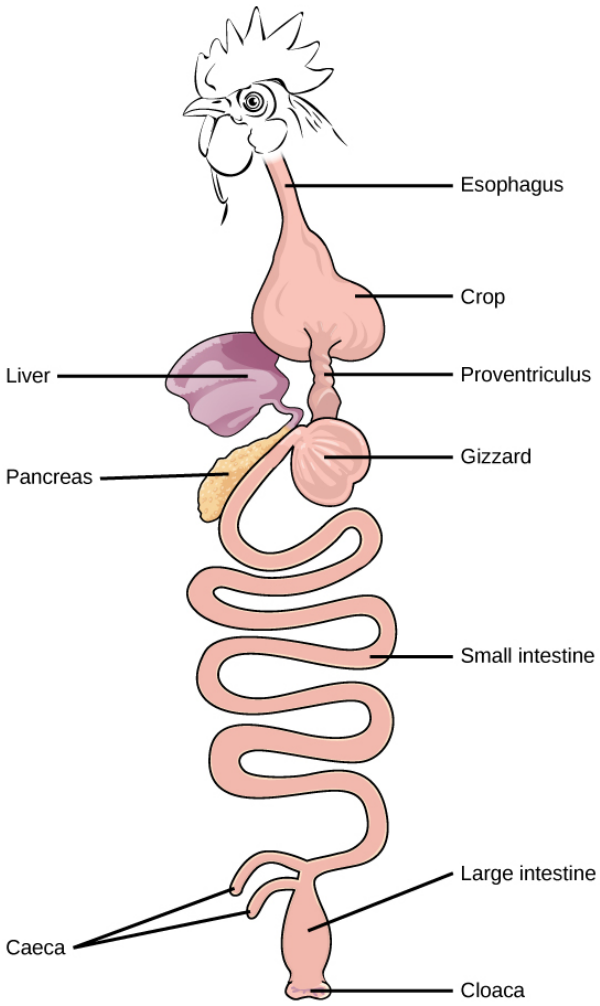


Figure 15.7. The avian esophagus has a pouch, called a crop, which stores food. Food passes from the crop to the first of two stomachs, called the proventriculus, which contains digestive juices that break down food. From the proventriculus, the food enters the second stomach, called the gizzard, which grinds food. Some birds swallow stones or grit, which are stored in the gizzard, to aid the grinding process. Birds do not have separate openings to excrete urine and feces. Instead, uric acid from the kidneys is secreted into the large intestine and combined with waste from the digestive process. This waste is excreted through an opening called the cloaca.

Parts of the Digestive System

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

Oral Cavity

The oral cavity, or mouth, is the point of entry of food into the digestive system, illustrated in Figure 15.9. The food consumed is broken into smaller particles by mastication, the chewing action of the teeth. All mammals have teeth and can chew their food.

The extensive chemical process of digestion begins in the mouth. As food is being chewed, saliva, produced by the salivary glands, mixes with the food. Saliva is a watery substance produced in the mouths of many animals. There are three major glands that secrete saliva—the parotid, the submandibular, and the sublingual. Saliva contains mucus that moistens food and buffers the pH of the food. Saliva also contains immunoglobulins and lysozymes, which have 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. Another enzyme called **lipase** is produced by the cells in the tongue. Lipases are a class of enzymes that can break down triglycerides. The lingual lipase begins the breakdown of fat components in the food. The chewing and wetting action provided by the teeth and saliva prepare the food into a mass called the **bolus** for swallowing. The tongue helps in swallowing—moving the bolus from the mouth into the pharynx. The pharynx opens to two passageways: the trachea, which leads to the lungs, and the esophagus, which leads to the stomach. The trachea has an opening called the glottis, which is covered by a cartilaginous flap called the epiglottis. When swallowing, the epiglottis closes the glottis and food passes into the esophagus and not the trachea. This arrangement allows food to be kept out of the trachea.

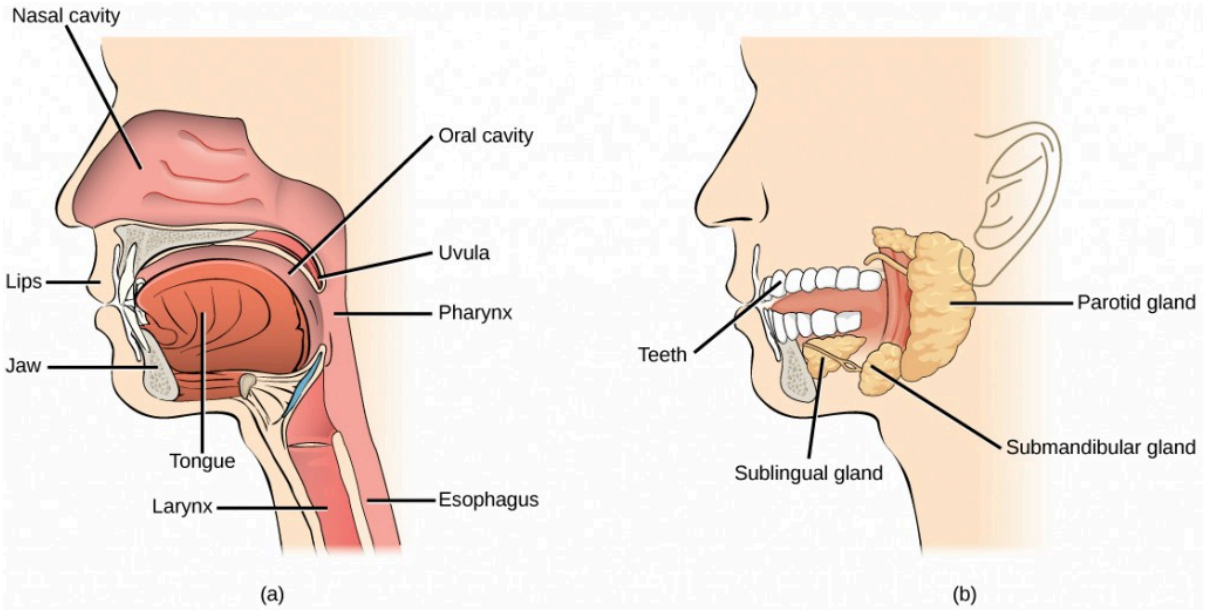


Figure 15.9.
Digestion of food begins in the (a) oral cavity. Food is masticated by teeth and moistened by saliva secreted from the (b) salivary glands. Enzymes in the saliva begin to digest starches and fats. With the help of the tongue, the resulting bolus is moved into the esophagus by swallowing. (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 15.10. The peristalsis wave is unidirectional—it moves food from the mouth to the stomach, and reverse movement is not possible. The peristaltic movement of the esophagus is an involuntary reflex; it takes place in response to the act of swallowing.

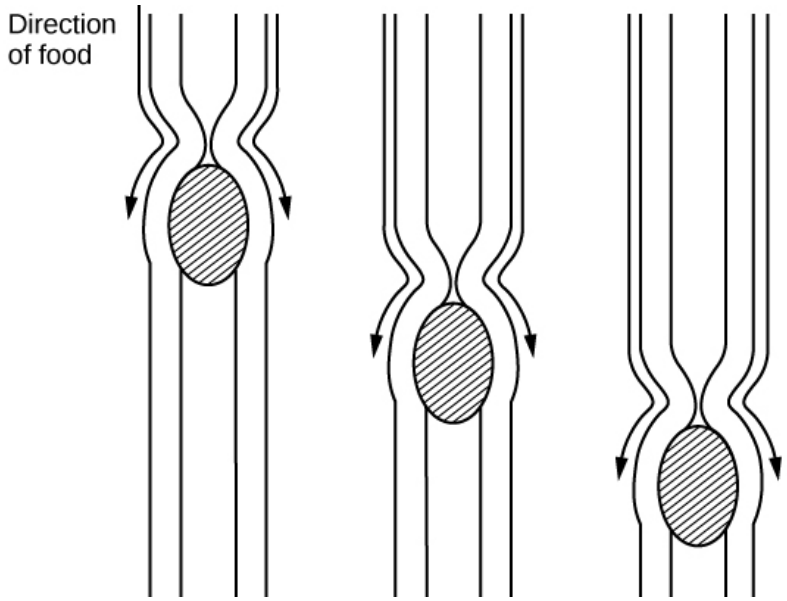


Figure 15.10. The esophagus transfers food from the mouth to the stomach through peristaltic movements.

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 15.11. 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.

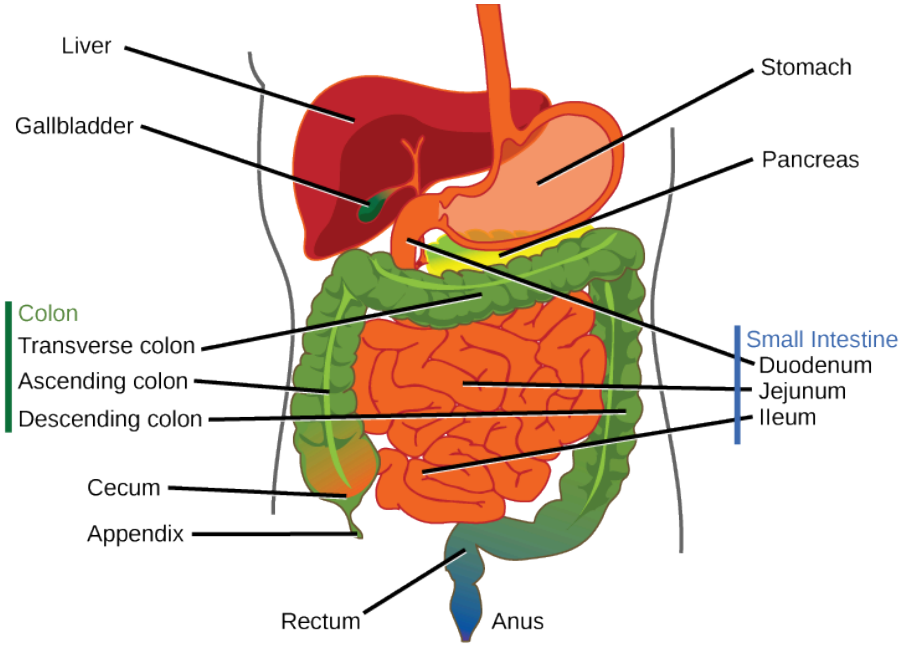


Figure 15.11. The human stomach has an extremely acidic environment where most of the protein gets digested. (credit: modification of work by Mariana Ruiz Villareal)

Which of the following statements about the digestive system is false?

1. Chyme is a mixture of food and digestive juices that is produced in the stomach.
2. Food enters the large intestine before the small intestine.
3. In the small intestine, chyme mixes with bile, which emulsifies fats.
4. The stomach is separated from the small intestine by the pyloric sphincter.

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

previously mentioned, the enzyme pepsin is synthesized in the inactive form. This protects the chief cells, because pepsinogen does not have the same enzyme functionality of pepsin. Second, the stomach has a thick mucus lining that protects the underlying tissue from the action of the digestive juices. When this mucus lining is ruptured, ulcers can form in the stomach. Ulcers are open wounds in or on an organ caused by bacteria **(Helicobacter pylori)** when the mucus lining is ruptured and fails to reform.

Small Intestine

Chyme moves from the stomach to the small intestine. The **small intestine** is the organ where the digestion of protein, fats, and carbohydrates is completed. The small intestine is a long tube-like organ with a highly folded surface containing finger-like projections called the **villi**. The apical surface of each villus has many microscopic projections called microvilli. These structures, illustrated in Figure 15.12, are lined with epithelial cells on the luminal side and allow for the nutrients to be absorbed from the digested food and absorbed into the blood stream on the other side. The villi and microvilli, with their many folds, increase the surface area of the intestine and increase absorption efficiency of the nutrients. Absorbed nutrients in the blood are carried into the hepatic portal vein, which leads to the liver. There, the liver regulates the distribution of nutrients to the rest of the body and removes toxic substances, including drugs, alcohol, and some pathogens.

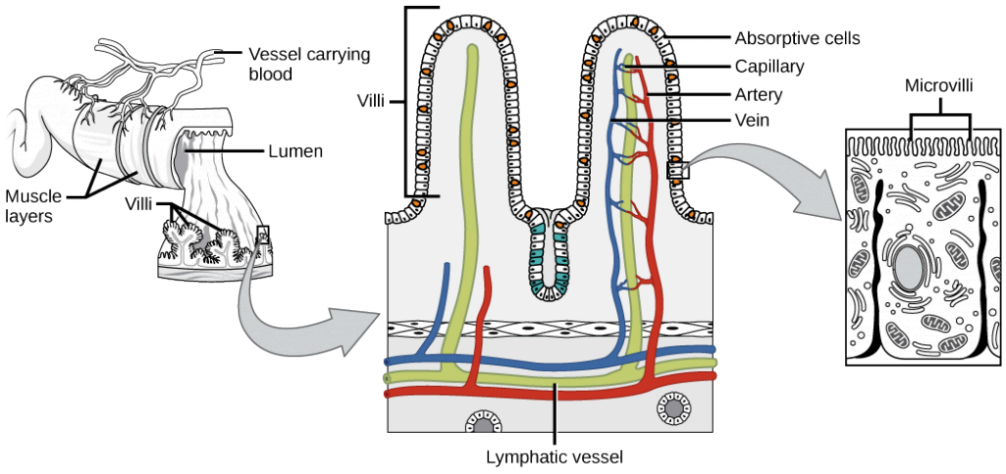


Figure 15.12. Villi are folds on the small intestine lining that increase the surface area to facilitate the absorption of nutrients.

Which of the following statements about the small intestine is false?

1. Absorptive cells that line the small intestine have microvilli, small projections that increase surface area and aid in the absorption of food.
2. The inside of the small intestine has many folds, called villi.
3. Microvilli are lined with blood vessels as well as lymphatic vessels.
4. The inside of the small intestine is called the lumen.

The human small intestine is over 6m long and is divided into three parts: the duodenum, the jejunum, and the ileum. The “C-shaped,” fixed part of the small intestine is called the **duodenum** and is shown

in Figure 15.11. The duodenum is separated from the stomach by the pyloric sphincter which opens to allow chyme to move from the stomach to the duodenum. In the duodenum, chyme is mixed with pancreatic juices in an alkaline solution rich in bicarbonate that neutralizes the acidity of chyme and acts as a buffer. Pancreatic juices also contain several digestive enzymes. Digestive juices from the pancreas, liver, and gallbladder, as well as from gland cells of the intestinal wall itself, enter the duodenum. **Bile** is produced in the liver and stored and concentrated in the gallbladder. Bile contains bile salts which emulsify lipids while the pancreas produces enzymes that catabolize starches, disaccharides, proteins, and fats. These digestive juices break down the food particles in the chyme into glucose, triglycerides, and amino acids. Some chemical digestion of food takes place in the duodenum. Absorption of fatty acids also takes place in the duodenum.

The second part of the small intestine is called the **jejunum**, shown in Figure 15.11. 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 15.11 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 15.13, reabsorbs the water from the undigested food material and processes the waste material. The human large intestine is much smaller in length compared to the small intestine but larger in diameter. It has three parts: the cecum, the colon, and the rectum. The cecum joins the ileum to the colon and is the receiving pouch for the waste matter. The colon is home to many bacteria or “intestinal flora” that aid in the digestive processes. The colon can be divided into four regions, the ascending colon, the transverse colon, the descending colon and the sigmoid colon. The main functions of the colon are to extract the water and mineral salts from undigested food, and to store waste material. Carnivorous mammals have a shorter large intestine compared to herbivorous mammals due to their diet.

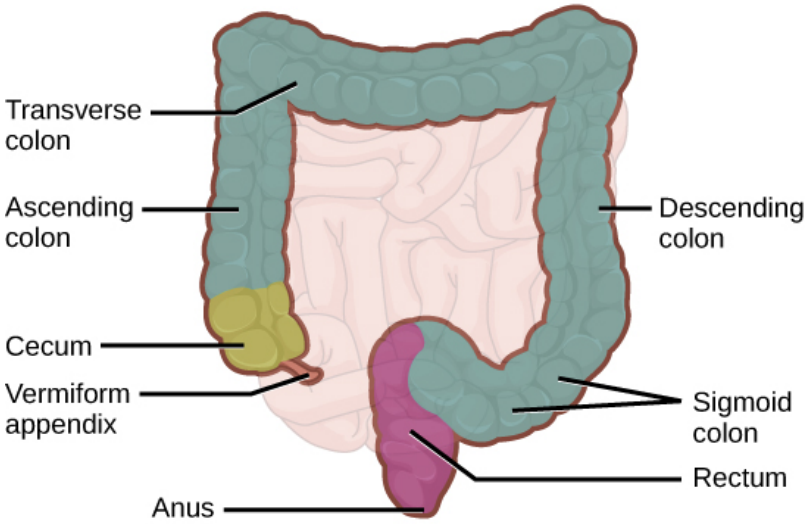


Figure 15.13.
The large intestine reabsorbs water from undigested food and stores waste material until it is eliminated.

Rectum and Anus

The rectum is the terminal end of the large intestine, as shown in Figure 15.13. 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.

Summary

Different animals have evolved different types of digestive systems specialized to meet their dietary needs. Humans and many other animals have monogastric digestive systems with a single-chambered stomach. Birds have evolved a digestive system that includes a gizzard where the food is crushed into smaller pieces. This compensates for their inability to masticate. Ruminants that consume large amounts of plant material have a multi-chambered stomach that digests roughage. Pseudo-ruminants have similar digestive processes as ruminants but do not have the four-compartment stomach. Processing food involves ingestion (eating), digestion (mechanical and enzymatic breakdown of large molecules), absorption (cellular uptake of nutrients), and elimination (removal of undigested waste as feces).

Many organs work together to digest food and absorb nutrients. The mouth is the point of ingestion and the location where both mechanical and chemical breakdown of food begins. Saliva contains an enzyme called amylase that breaks down carbohydrates. The food bolus travels through the esophagus by peristaltic movements to the stomach. The stomach has an extremely acidic environment. An enzyme called pepsin digests protein in the stomach. Further digestion and absorption take place in the small intestine. The large intestine reabsorbs water from the undigested food and stores waste until elimination.

Exercises

1. Which of the following statements about the digestive system is false?

    1. Chyme is a mixture of food and digestive juices that is produced in the stomach.
    2. Food enters the large intestine before the small intestine.
    3. In the small intestine, chyme mixes with bile, which emulsifies fats.
    4. The stomach is separated from the small intestine by the pyloric sphincter.

2. Which of the following statements about the small intestine is false?

    1. Absorptive cells that line the small intestine have microvilli, small projections that increase surface area and aid in the absorption of food.
    2. The inside of the small intestine has many folds, called villi.
    3. Microvilli are lined with blood vessels as well as lymphatic vessels.
    4. The inside of the small intestine is called the lumen.

3. Which of the following is a pseudo-ruminant?

    1. cow
    2. pig
    3. crow
    4. horse

4. Which of the following statements is untrue?

    1. Roughage takes a long time to digest.
    2. Birds eat large quantities at one time so that they can fly long distances.
    3. Cows do not have upper teeth.
    4. In pseudo-ruminants, roughage is digested in the cecum.

5. The acidic nature of chyme is neutralized by _____.

    1. potassium hydroxide
    2. sodium hydroxide
    3. bicarbonates
    4. vinegar

6. The digestive juices from the liver are delivered to the _____.

    1. stomach
    2. liver
    3. duodenum
    4. colon

7. How does the polygastric digestive system aid in digesting roughage?

8. How do birds digest their food in the absence of teeth?
9. What is the role of the accessory organs in digestion?
10. Explain how the villi and microvilli aid in absorption.

Answers

1. B
2. C
3. D
4. B
5. C
6. C
7. Animals with a polygastric digestive system have a multi-chambered stomach. The four compartments of the stomach are called the rumen, reticulum, omasum, and abomasum. These chambers contain many microbes that break down the cellulose and ferment the ingested food. The abomasum is the "true" stomach and is the equivalent of a monogastric stomach chamber where gastric juices are secreted. The four-compartment gastric chamber provides larger space and the microbial support necessary for ruminants to digest plant material.
8. Birds have a stomach chamber called a gizzard. Here, the food is stored, soaked, and ground into finer particles, often using pebbles. Once this process is complete, the digestive juices take over in the proventriculus and continue the digestive process.
9. Accessory organs play an important role in producing and delivering digestive juices to the intestine during digestion and absorption. Specifically, the salivary glands, liver, pancreas, and gallbladder play important roles. Malfunction of any of these organs can lead to disease states.
10. The villi and microvilli are folds on the surface of the small intestine. These folds increase the surface area of the intestine and provide more area for the absorption of nutrients.

Glossary

alimentary canal: tubular digestive system with a mouth and anus
anus: exit point for waste material
bile: digestive juice produced by the liver; important for digestion of lipids
bolus: mass of food resulting from chewing action and wetting by saliva
carnivore: animal that consumes animal flesh
chyme: mixture of partially digested food and stomach juices
digestion: mechanical and chemical break down of food into small organic fragments
duodenum: first part of the small intestine where a large part of digestion of carbohydrates and fats occurs
endocrine system: system that controls the response of the various glands in the body and the release of hormones at the appropriate times
esophagus: tubular organ that connects the mouth to the stomach
essential nutrient: nutrient that cannot be synthesized by the body; it must be obtained from food
gallbladder: organ that stores and concentrates bile

gastric inhibitory peptide: hormone secreted by the small intestine in the presence of fatty acids and sugars; it also inhibits acid production and peristalsis in order to slow down the rate at which food enters the small intestine
gastrin: hormone which stimulates hydrochloric acid secretion in the stomach
gastrovascular cavity: digestive system consisting of a single opening
gizzard: muscular organ that grinds food
herbivore: animal that consumes strictly plant diet
ileum: last part of the small intestine; connects the small intestine to the large intestine; important for absorption of B-12
ingestion: act of taking in food
jejunum: second part of the small intestine
lactase: enzyme that breaks down lactose into glucose and galactose
large intestine: digestive system organ that reabsorbs water from undigested material and processes waste matter
lipase: enzyme that chemically breaks down lipids
liver: organ that produces bile for digestion and processes vitamins and lipids
maltase: enzyme that breaks down maltose into glucose
mineral: inorganic, elemental molecule that carries out important roles in the body
monogastric: digestive system that consists of a single-chambered stomach
omnivore: animal that consumes both plants and animals
pancreas: gland that secretes digestive juices
pepsinogen: inactive form of pepsin
pepsin: enzyme found in the stomach whose main role is protein digestion
peristalsis: wave-like movements of muscle tissue
proventriculus: glandular part of a bird's stomach
rectum: area of the body where feces is stored until elimination
roughage: component of food that is low in energy and high in fiber
ruminant: animal with a stomach divided into four compartments
salivary amylase: enzyme found in saliva, which converts carbohydrates to maltose
small intestine: organ where digestion of protein, fats, and carbohydrates is completed
somatostatin: hormone released to stop acid secretion when the stomach is empty
sphincter: band of muscle that controls movement of materials throughout the digestive tract
stomach: saclike organ containing acidic digestive juices
villi: folds on the inner surface of the small intestine whose role is to increase absorption area
vitamin: organic substance necessary in small amounts to sustain life

15.2 Nutrition and Energy Production

Learning Objectives

By the end of this section, you will be able to:

- Explain why an animal's diet should be balanced and meet the needs of the body
- Define the primary components of food
- Describe the essential nutrients required for cellular function that cannot be synthesized by the animal body
- Explain how energy is produced through diet and digestion
- Describe how excess carbohydrates and energy are stored in the body

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 15.14

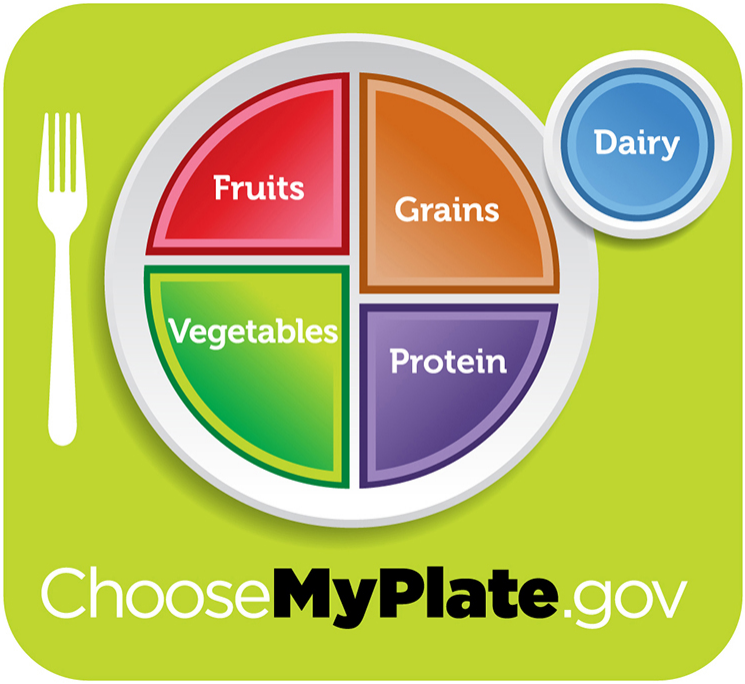


Figure 15.14.
For humans, a balanced diet includes fruits, vegetables, grains, and protein. (credit: USDA)

Concept in Action

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

Organic Precursors

The organic molecules required for building cellular material and tissues must come from food. Carbohydrates or sugars are the primary 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

An interactive or media element has been excluded from this version of the text. You can view it online here: https://opentextbc.ca/biology/?p=4816

https://www.youtube.com/watch?v=sk6iAl4qyEA&list=PL50LJVchZ8-JScvWocCqyrer9AznISngh&index=4

An interactive or media element has been excluded from this version of the text. You can view it online here: https://opentextbc.ca/biology/?p=4816

https://www.youtube.com/watch?v=zMArRi0hqqw&index=10&list=PL50LJVchZ8-JScvWocCqyrer9AznISngh

An interactive or media element has been excluded from this version of the text. You can view it online here: https://opentextbc.ca/biology/?p=4816

https://www.youtube.com/watch?v=s8I78Vn5sHM&index=8&list=PL50LJVchZ8-JScvWocCqyrer9AznISngh

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 15.1 and Table 15.2. Both fat-soluble and water-soluble vitamins must be obtained from food. **Minerals,** listed in Table 15.3, 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 15.4.

Table 15.1.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 15.2. 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

Figure 15.15. A healthy diet should include a variety of foods to ensure that needs for essential nutrients are met. (credit: Keith Weller, USDA ARS)

Table 15.3. 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
*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

Mineral	Function	Deficiencies Can Lead To	Sources
*Greater than 200mg/day required			

Table 15.4. 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.	

Food Energy and ATP

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

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

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

ATP is required for all cellular functions. It is used to build the organic molecules that are required for cells and tissues; it provides energy for muscle contraction and for the transmission of electrical signals in the nervous system. When the amount of ATP is available in excess of the body's requirements, the liver uses the excess ATP and excess glucose to produce molecules called glycogen. Glycogen is a polymeric form of glucose and is stored in the liver and skeletal muscle cells. When blood sugar drops, the liver releases glucose from stores of glycogen. Skeletal muscle converts glycogen to glucose during intense exercise. The process of converting glucose and excess ATP to glycogen and the storage of excess energy is an evolutionarily important step in helping animals deal with mobility, food shortages, and famine.

Obesity

Obesity is a major health concern in the United States, and there is a growing focus on reducing obesity and the diseases it may lead to, such as type-2 diabetes, cancers of the colon and breast, and cardiovascular disease. How does the food consumed contribute to obesity?

Fatty foods are calorie-dense, meaning that they have more calories per unit mass than carbohydrates or proteins. One gram of carbohydrates has four calories, one gram of protein has four calories, and one gram of fat has nine calories. Animals tend to seek lipid-rich food for their higher energy content.

The signals of hunger ("time to eat") and satiety ("time to stop eating") are controlled in the hypothalamus region of the brain. Foods that are rich in fatty acids tend to promote satiety more than foods that are rich only in carbohydrates.

Excess carbohydrate and ATP are used by the liver to synthesize glycogen. The pyruvate produced during glycolysis is used to synthesize fatty acids. When there is more glucose in the body than required, the resulting excess pyruvate is converted into molecules that eventually result in the synthesis of fatty acids within the body. These fatty acids are stored in adipose cells—the fat cells in the mammalian body whose primary role is to store fat for later use.

It is important to note that some animals benefit from obesity. Polar bears and seals need body fat for insulation and to keep them from losing body heat during Arctic winters. When food is scarce, stored body fat provides energy for maintaining homeostasis. Fats prevent famine in mammals, allowing them to access energy when food is not available on a daily basis; fats are stored when a large kill is made or lots of food is available.

Summary

Animal diet should be balanced and meet the needs of the body. Carbohydrates, proteins, and fats are the primary components of food. Some essential nutrients are required for cellular function but cannot be produced by the animal body. These include vitamins, minerals, some fatty acids, and some amino acids. Food intake in more than necessary amounts is stored as glycogen in the liver and muscle cells, and in fat cells. Excess adipose storage can lead to obesity and serious health problems. ATP is the energy currency of the cell and is obtained from the metabolic pathways. Excess carbohydrates and energy are stored as glycogen in the body.

Exercises

1. Which of the following statements is not true?
    1. Essential nutrients can be synthesized by the body.
    2. Vitamins are required in small quantities for bodily function.
    3. Some amino acids can be synthesized by the body, while others need to be obtained from diet.
    4. Vitamins come in two categories: fat-soluble and water-soluble.
2. Which of the following is a water-soluble vitamin?
    1. vitamin A
    2. vitamin E
    3. vitamin K
    4. vitamin C
3. What is the primary fuel for the body?
    1. carbohydrates
    2. lipids
    3. protein
    4. glycogen
4. Excess glucose is stored as _____.
    1. fat
    2. glucagon
    3. glycogen
    4. it is not stored in the body
5. What are essential nutrients?
6. What is the role of minerals in maintaining good health?
7. Discuss why obesity is a growing epidemic.
8. There are several nations where malnourishment is a common occurrence. What may be some of the health challenges posed by malnutrition?

Answers

1. A
2. D
3. A
4. C
5. Essential nutrients are those nutrients that must be obtained from the diet because they cannot be

produced by the body. Vitamins and minerals are examples of essential nutrients.

6. Minerals—such as potassium, sodium, and calcium—are required for the functioning of many cellular processes, including muscle contraction and nerve conduction. While minerals are required in trace amounts, not having minerals in the diet can be potentially harmful.
7. In the United States, obesity, particularly childhood obesity, is a growing concern. Some of the contributors to this situation include sedentary lifestyles and consuming more processed foods and less fruits and vegetables. As a result, even young children who are obese can face health concerns.
8. Malnutrition, often in the form of not getting enough calories or not enough of the essential nutrients, can have severe consequences. Many malnourished children have vision and dental problems, and over the years may develop many serious health problems.

Glossary

bile: digestive juice produced by the liver; important for digestion of lipids

carboxypeptidase: protease that breaks down peptides to single amino acids; secreted by the brush border of the small intestine

chyme: mixture of partially digested food and stomach juices

digestion: mechanical and chemical break down of food into small organic fragments

essential nutrient: nutrient that cannot be synthesized by the body; it must be obtained from food

large intestine: digestive system organ that reabsorbs water from undigested material and processes waste matter

liver: organ that produces bile for digestion and processes vitamins and lipids

mineral: inorganic, elemental molecule that carries out important roles in the body

small intestine: organ where digestion of protein, fats, and carbohydrates is completed

stomach: sac-like organ containing acidic digestive juices

vitamin: organic substance necessary in small amounts to sustain life

15.3 Digestive System Processes

Learning Objectives

By the end of this section, you will be able to:

- Describe the process of digestion
- Detail the steps involved in digestion and absorption
- Define elimination
- Explain the role of both the small and large intestines in absorption

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

Ingestion

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

Digestion and Absorption

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

Carbohydrates

The digestion of carbohydrates begins in the mouth. The salivary enzyme amylase begins the breakdown of food starches into maltose, a disaccharide. As the bolus of food travels through the

esophagus to the stomach, no significant digestion of carbohydrates takes place. The esophagus produces no digestive enzymes but does produce mucous for lubrication. The acidic environment in the stomach stops the action of the amylase enzyme.

The next step of carbohydrate digestion takes place in the duodenum. Recall that the chyme from the stomach enters the duodenum and mixes with the digestive secretion from the pancreas, liver, and gallbladder. Pancreatic juices also contain amylase, which continues the breakdown of starch and glycogen into maltose, a disaccharide. The disaccharides are broken down into monosaccharides by enzymes called **maltases**

, sucrases, and **lactases**, which are also present in the brush border of the small intestinal wall. Maltase breaks down maltose into glucose. Other disaccharides, such as sucrose and lactose are broken down by sucrase and lactase, respectively. Sucrase breaks down sucrose (or “table sugar”) into glucose and fructose, and lactase breaks down lactose (or “milk sugar”) into glucose and galactose. The monosaccharides (glucose) thus produced are absorbed and then can be used in metabolic pathways to harness energy. The monosaccharides are transported across the intestinal epithelium into the bloodstream to be transported to the different cells in the body. The steps in carbohydrate digestion are summarized in Figure 15.16 and Table 15.5.

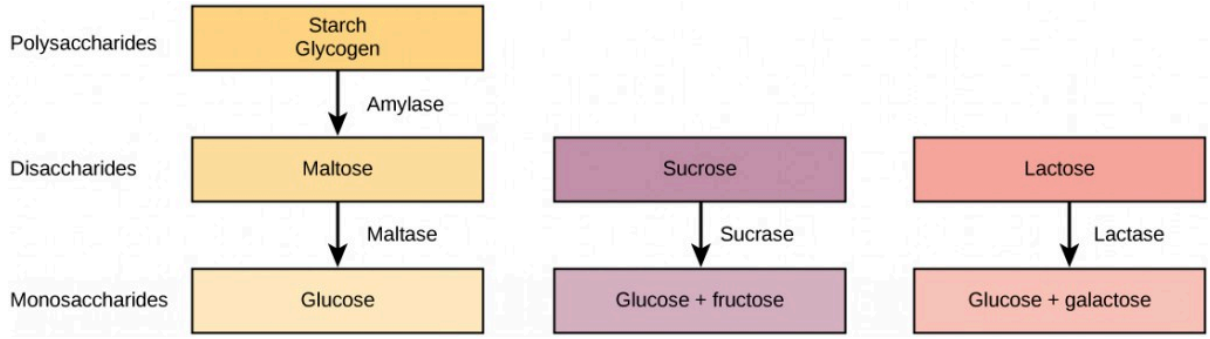


Figure 15.16. Digestion of carbohydrates is performed by several enzymes. Starch and glycogen are broken down into glucose by amylase and maltase. Sucrose (table sugar) and lactose (milk sugar) are broken down by sucrase and lactase, respectively.

Table15.5 Digestion of Carbohydrates

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

Protein

A large part of protein digestion takes place in the stomach. The enzyme pepsin plays an important role

in the digestion of proteins by breaking down the intact protein to peptides, which are short chains of four to nine amino acids. In the duodenum, other enzymes—**trypsin, elastase**, and **chymotrypsin**—act on the peptides reducing them to smaller peptides. Trypsin elastase, carboxypeptidase, and chymotrypsin are produced by the pancreas and released into the duodenum where they act on the chyme. 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 15.17 and Table 15.6.

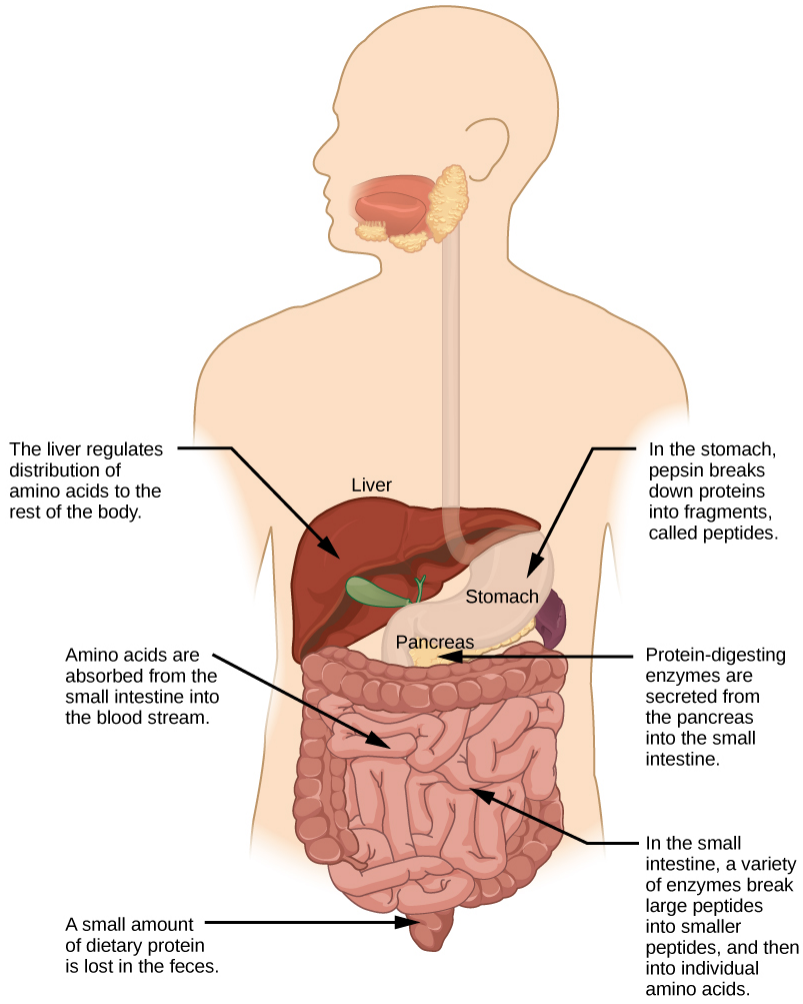


Figure 15.17
Protein digestion is a multistep process that begins in the stomach and continues through the intestines.

Table 15.6. 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

Lipids

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

Why is emulsification important for digestion of lipids? Pancreatic juices contain enzymes called lipases (enzymes that break down lipids). If the lipid in the chyme aggregates into large globules, very little surface area of the lipids is available for the lipases to act on, leaving lipid digestion incomplete. By forming an emulsion, bile salts increase the available surface area of the lipids many fold. The pancreatic lipases can then act on the lipids more efficiently and digest them, as detailed in Figure 15.18. Lipases break down the lipids into fatty acids and glycerides. These molecules can pass through the plasma membrane of the cell and enter the epithelial cells of the intestinal lining. The bile salts surround long-chain fatty acids and monoglycerides forming tiny spheres called micelles. The micelles move into the brush border of the small intestine absorptive cells where the long-chain fatty acids and monoglycerides diffuse out of the micelles into the absorptive cells leaving the micelles behind in the chyme. The long-chain fatty acids and monoglycerides recombine in the absorptive cells to form triglycerides, which aggregate into globules and become coated with proteins. These large spheres are called **chylomicrons**. Chylomicrons contain triglycerides, cholesterol, and other lipids and have proteins on their surface. The surface is also composed of the hydrophilic phosphate “heads” of

phospholipids. Together, they enable the chylomicron to move in an aqueous environment without exposing the lipids to water. Chylomicrons leave the absorptive cells via exocytosis. Chylomicrons enter the lymphatic vessels, and then enter the blood in the subclavian vein.

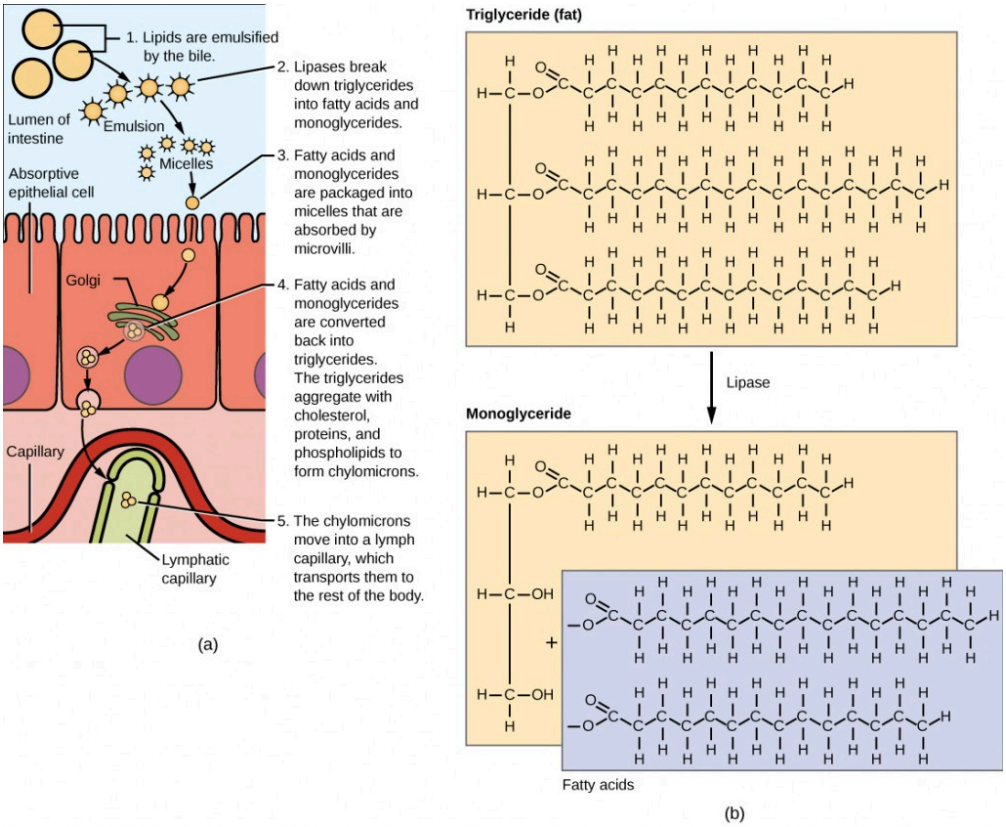


Figure 15.18.
Lipids are digested and absorbed in the small intestine.

Vitamins

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

Concept in Action

This website has an overview of the digestion of protein, fat, and carbohydrates.

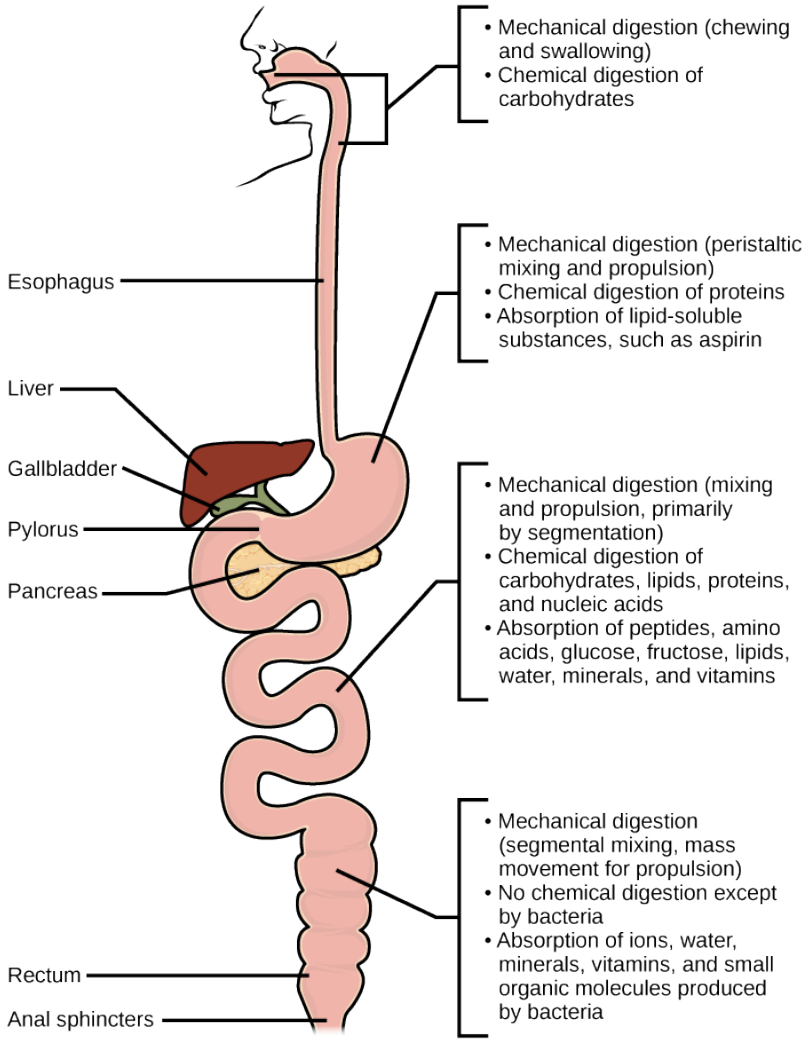


Figure 15.19. Mechanical and chemical digestion of food takes place in many steps, beginning in the mouth and ending in the rectum.

Which of the following statements about digestive processes is true?

1. Amylase, maltase, and lactase in the mouth digest carbohydrates.
2. Trypsin and lipase in the stomach digest protein.
3. Bile emulsifies lipids in the small intestine.
4. No food is absorbed until the small intestine.

Elimination

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

set up the urge to eliminate. The solid waste is eliminated through the anus using peristaltic movements of the rectum.

Common Problems with Elimination

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

Emesis

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

Summary

Animal diet should be balanced and meet the needs of the body. Carbohydrates, proteins, and fats are the primary components of food. Some essential nutrients are required for cellular function but cannot be produced by the animal body. These include vitamins, minerals, some fatty acids, and some amino acids. Food intake in more than necessary amounts is stored as glycogen in the liver and muscle cells, and in fat cells. Excess adipose storage can lead to obesity and serious health problems. ATP is the energy currency of the cell and is obtained from the metabolic pathways. Excess carbohydrates and energy are stored as glycogen in the body.

Exercises

1. Where does the majority of protein digestion take place?
    1. stomach
    2. duodenum
    3. mouth
    4. jejunum
2. Lipases are enzymes that break down _____.
    1. disaccharides
    2. lipids
    3. proteins
    4. cellulose

3. Explain why some dietary lipid is a necessary part of a balanced diet.

Answers

1. A
2. B
3. Lipids add flavor to food and promote a sense of satiety or fullness. Fatty foods are sources of high energy; one gram of lipid contains nine calories. Lipids are also required in the diet to aid the absorption of lipid-soluble vitamins and for the production of lipid-soluble hormones.

Glossary

aminopeptidase: protease that breaks down peptides to single amino acids; secreted by the brush border of small intestine
anus: exit point for waste material
bile: digestive juice produced by the liver; important for digestion of lipids
bolus: mass of food resulting from chewing action and wetting by saliva
carboxypeptidase: protease that breaks down peptides to single amino acids; secreted by the brush border of the small intestine
chylomicron: small lipid globule
chyme: mixture of partially digested food and stomach juices
chymotrypsin: pancreatic protease
digestion: mechanical and chemical break down of food into small organic fragments
dipeptidase: protease that breaks down peptides to single amino acids; secreted by the brush border of small intestine
duodenum: first part of the small intestine where a large part of digestion of carbohydrates and fats occurs
elastase: pancreatic protease
esophagus: tubular organ that connects the mouth to the stomach
essential nutrient: nutrient that cannot be synthesized by the body; it must be obtained from food
gallbladder: organ that stores and concentrates bile
ingestion: act of taking in food
jejunum: second part of the small intestine
lactase: enzyme that breaks down lactose into glucose and galactose
large intestine: digestive system organ that reabsorbs water from undigested material and processes waste matter
lipase: enzyme that chemically breaks down lipids
liver: organ that produces bile for digestion and processes vitamins and lipids
maltase: enzyme that breaks down maltose into glucose
mineral: inorganic, elemental molecule that carries out important roles in the body
pancreas: gland that secretes digestive juices
pepsin: enzyme found in the stomach whose main role is protein digestion
rectum: area of the body where feces is stored until elimination
small intestine: organ where digestion of protein, fats, and carbohydrates is completed
stomach: sac-like organ containing acidic digestive juices
sucrase: enzyme that breaks down sucrose into glucose and fructose

trypsin: pancreatic protease that breaks down protein
vitamin: organic substance necessary in small amounts to sustain life

15.4 Digestive System Regulation

Learning Objectives

By the end of this section, you will be able to:

- Discuss the role of neural regulation in digestive processes
- Explain how hormones regulate digestion

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 15.20, 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 15.20.
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 phas**, 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 gastrin 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. Gastrin stimulates the release of stomach acid, or hydrochloric acid (HCl) which aids in the digestion of the proteins. However, when the stomach is emptied, the acidic environment need not be maintained and a hormone called **somatostatin** stops the release of hydrochloric acid. This is controlled by a negative feedback mechanism.

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.

Concept in Action

Visit

this website to learn more about the endocrine system. Review the text and watch the animation of how control is implemented in the endocrine system.

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.

Summary

The brain and the endocrine system control digestive processes. The brain controls the responses of

hunger and satiety. The endocrine system controls the release of hormones and enzymes required for digestion of food in the digestive tract.

Exercises

1. Which of the following is a pseudo-ruminant?
    1. cow
    2. pig
    3. crow
    4. horse
2. Which of the following statements is untrue?
    1. Roughage takes a long time to digest.
    2. Birds eat large quantities at one time so that they can fly long distances.
    3. Cows do not have upper teeth.
    4. In pseudo-ruminants, roughage is digested in the cecum.
3. The acidic nature of chyme is neutralized by _____.
    1. potassium hydroxide
    2. sodium hydroxide
    3. bicarbonates
    4. vinegar
4. The digestive juices from the liver are delivered to the _____.
    1. stomach
    2. liver
    3. duodenum
    4. colon
5. Which of the following statements is not true?
    1. Essential nutrients can be synthesized by the body.
    2. Vitamins are required in small quantities for bodily function.
    3. Some amino acids can be synthesized by the body, while others need to be obtained from diet.
    4. Vitamins come in two categories: fat-soluble and water-soluble.
6. Which of the following is a water-soluble vitamin?
    1. vitamin A
    2. vitamin E

3. vitamin K
4. vitamin C

7. What is the primary fuel for the body?

    1. carbohydrates
    2. lipids
    3. protein
    4. glycogen

8. Excess glucose is stored as _____.

    1. fat
    2. glucagon
    3. glycogen
    4. it is not stored in the body

9. Where does the majority of protein digestion take place?

    1. stomach
    2. duodenum
    3. mouth
    4. jejunum

10. Lipases are enzymes that break down _____.

    1. disaccharides
    2. lipids
    3. proteins
    4. cellulose

11. Which hormone controls the release of bile from the gallbladder

    1. pepsin
    2. amylase
    3. CCK
    4. gastrin

12. Which hormone stops acid secretion in the stomach?

    1. gastrin
    2. somatostatin
    3. gastric inhibitory peptide
    4. CCK

13. Describe how hormones regulate digestion.
14. Describe one or more scenarios where loss of hormonal regulation of digestion can lead to diseases.

Answers

1. D
2. B
3. C
4. C
5. A
6. D
7. A
8. C
9. A
10. B
11. C
12. B
13. Hormones control the different digestive enzymes that are secreted in the stomach and the intestine during the process of digestion and absorption. For example, the hormone gastrin stimulates stomach acid secretion in response to food intake. The hormone somatostatin stops the release of stomach acid.
14. There are many cases where loss of hormonal regulation can lead to illnesses. For example, the bilirubin produced by the breakdown of red blood cells is converted to bile by the liver. When there is malfunction of this process, there is excess bilirubin in the blood and bile levels are low. As a result, the body struggles with dealing with fatty food. This is why a patient suffering from jaundice is asked to eat a diet with almost zero fat.

Glossary

bile
digestive juice produced by the liver; important for digestion of lipids

cephalic phase
first phase of digestion, controlled by the neural response to the stimulus provided by food

cholecystokinin
hormone that stimulates the contraction of the gallbladder to release bile

chyme
mixture of partially digested food and stomach juices

digestion
mechanical and chemical break down of food into small organic fragments

duodenum
first part of the small intestine where a large part of digestion of carbohydrates and fats occurs

endocrine system
system that controls the response of the various glands in the body and the release of hormones at the appropriate times

esophagus
tubular organ that connects the mouth to the stomach

gallbladder
organ that stores and concentrates bile

gastric inhibitory peptide
hormone secreted by the small intestine in the presence of fatty acids and sugars; it also inhibits acid production and peristalsis in order to slow down the rate at which food enters the small intestine

gastric phase
digestive phase beginning once food enters the stomach; gastric acids and enzymes process the ingested materials

gastrin
hormone which stimulates hydrochloric acid secretion in the stomach

ingestion
act of taking in food

intestinal phase
third digestive phase; begins when chyme enters the small intestine triggering digestive secretions and controlling the rate of gastric emptying

jejunum
second part of the small intestine

liver
organ that produces bile for digestion and processes vitamins and lipids

pancreas
gland that secretes digestive juices

pepsin
enzyme found in the stomach whose main role is protein digestion

peristalsis
wave-like movements of muscle tissue

roughage
component of food that is low in energy and high in fiber

ruminant
animal with a stomach divided into four compartments

secretin
hormone which stimulates sodium bicarbonate secretion in the small intestine

small intestine
organ where digestion of protein, fats, and carbohydrates is completed

somatostatin
hormone released to stop acid secretion when the stomach is empty

stomach
saclike organ containing acidic digestive juices

vitamin
organic substance necessary in small amounts to sustain life

Chapter 16. The Nervous System

Figure 16.1.
An athlete's nervous system is hard at work during the planning and execution of a movement as precise as a high jump. Parts of the nervous system are involved in determining how hard to push off and when to turn, as well as controlling the muscles throughout the body that make this complicated movement possible without knocking the bar down—all in just a few seconds. (credit: modification of work by Shane T. McCoy, U.S. Navy)

Introduction

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

16.1 Neurons and Glial Cells

Learning Objectives

By the end of this section, you will be able to:

- List and describe the functions of the structural components of a neuron
- List and describe the four main types of neurons
- Compare the functions of different types of glial cells

Nervous systems throughout the animal kingdom vary in structure and complexity, as illustrated by the variety of animals shown in Figure 16.2. 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.

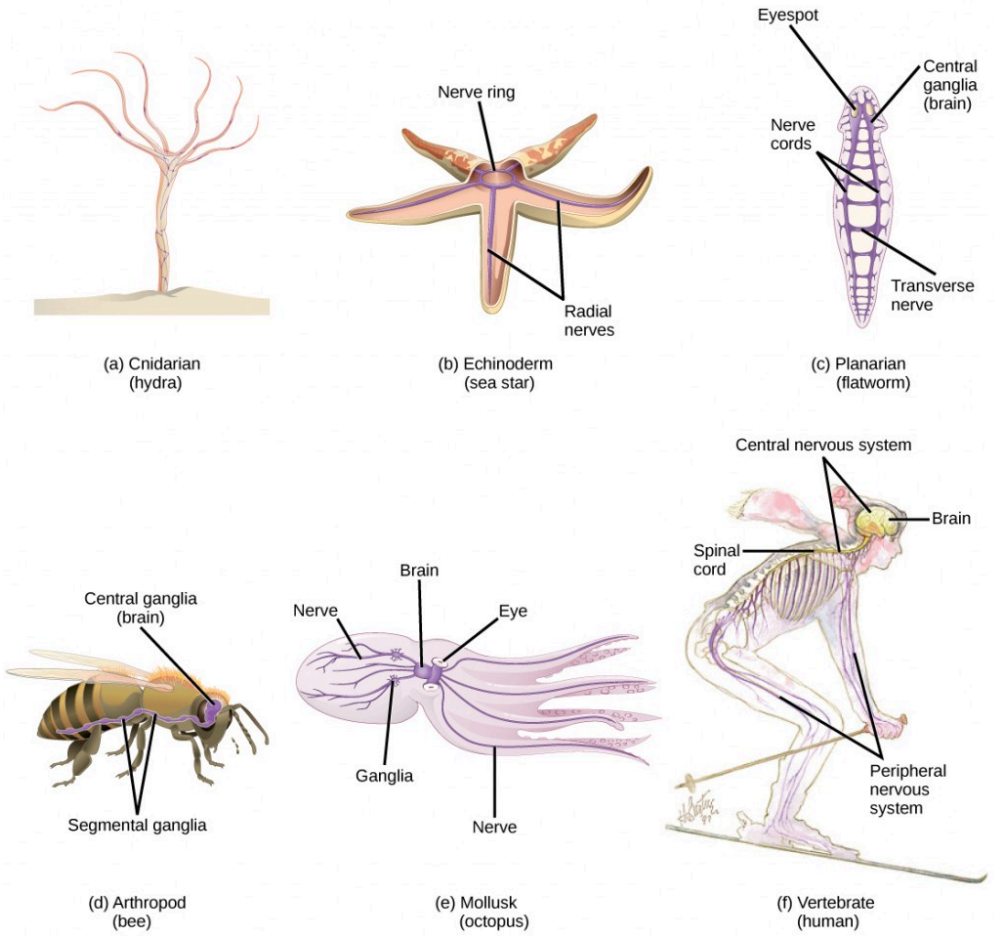


Figure 16.2. Nervous systems vary in structure and complexity. In (a) cnidarians, nerve cells form a decentralized nerve net. In (b) echinoderms, nerve cells are bundled into fibers called nerves. In animals exhibiting bilateral symmetry such as (c) planarians, neurons cluster into an anterior brain 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.

Concept in Action

Watch this video of biologist Mark Kirschner discussing the “flipping” phenomenon of vertebrate evolution.

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

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

Neurons

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

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

Parts of a Neuron

Like other cells, each neuron has a cell body (or soma) that contains a nucleus, smooth and rough endoplasmic reticulum, Golgi apparatus, mitochondria, and other cellular components. Neurons also contain unique structures, illustrated in Figure 16.3 for receiving and sending the electrical signals that make neuronal communication possible. **Dendrites** are tree-like structures that extend away from the cell body to receive messages from other neurons at specialized junctions called **synapses**. Although some neurons do not have any dendrites, some types of neurons have multiple dendrites. Dendrites can have small protrusions called dendritic spines, which further increase surface area for possible synaptic connections.

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

It is important to note that a single neuron does not act alone—neuronal communication depends on the connections that neurons make with one another (as well as with other cells, like muscle cells). Dendrites from a single neuron may receive synaptic contact from many other neurons. For example, dendrites from a Purkinje cell in the cerebellum are thought to receive contact from as many as 200,000 other neurons.

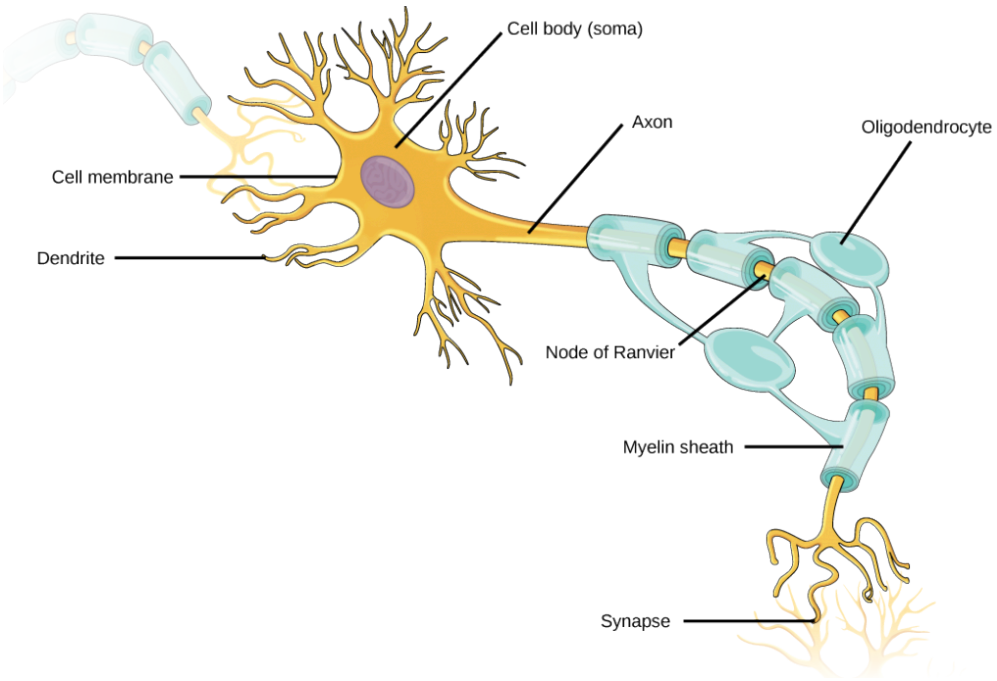


Figure 16.3. Neurons contain organelles common to many other cells, such as a nucleus and mitochondria. They also have more specialized structures, including dendrites and axons.

Which of the following statements is false?

1. The soma is the cell body of a nerve cell.
2. Myelin sheath provides an insulating layer to the dendrites.
3. Axons carry the signal from the soma to the target.

4. Dendrites carry the signal to the soma.

Types of Neurons

There are different types of neurons, and the functional role of a given neuron is intimately dependent on its structure. There is an amazing diversity of neuron shapes and sizes found in different parts of the nervous system (and across species), as illustrated by the neurons shown in Figure 16.4.

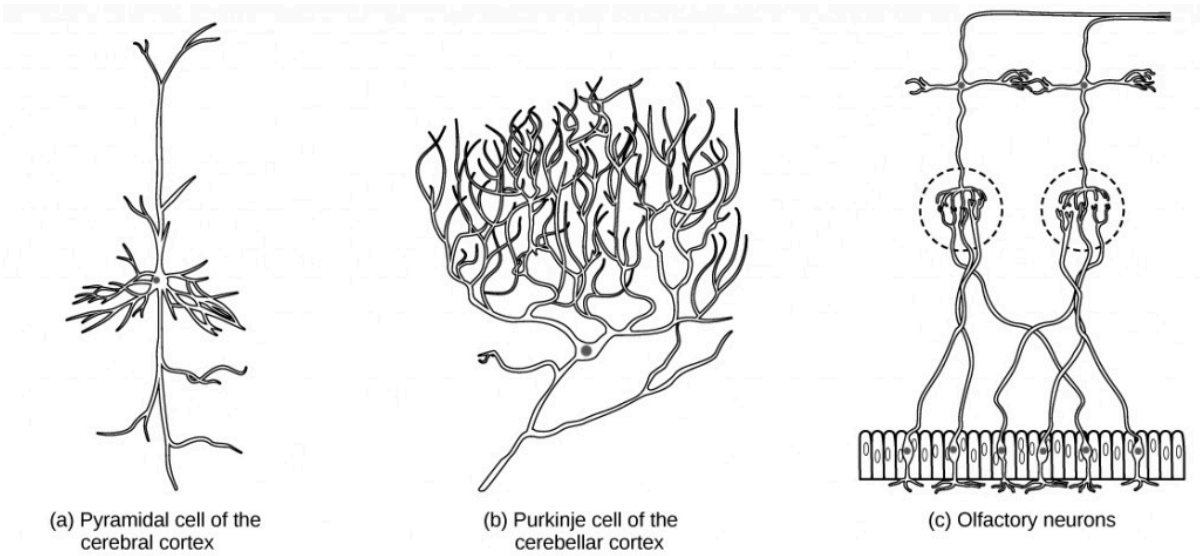


Figure 16.4. There is great diversity in the size and shape of neurons throughout the nervous system. Examples include (a) a pyramidal cell from the cerebral cortex, (b) a Purkinje cell from the cerebellar cortex, and (c) olfactory cells from the olfactory epithelium and olfactory bulb.

While there are many defined neuron cell subtypes, neurons are broadly divided into four basic types: unipolar, bipolar, multipolar, and pseudounipolar. Figure 16.5 illustrates these four basic neuron types. Unipolar neurons have only one structure that extends away from the soma. These neurons are not found in vertebrates but are found in insects where they stimulate muscles or glands. A bipolar neuron has one axon and one dendrite extending from the soma. An example of a bipolar neuron is a retinal bipolar cell, which receives signals from photoreceptor cells that are sensitive to light and transmits these signals to ganglion cells that carry the signal to the brain. Multipolar neurons are the most common type of neuron. Each multipolar neuron contains one axon and multiple dendrites. Multipolar neurons can be found in the central nervous system (brain and spinal cord). An example of a multipolar neuron is a Purkinje cell in the cerebellum, which has many branching dendrites but only one axon. Pseudounipolar cells share characteristics with both unipolar and bipolar cells. A pseudounipolar cell has a single process that extends from the soma, like a unipolar cell, but this process later branches into two distinct structures, like a bipolar cell. Most sensory neurons are pseudounipolar and have an axon that branches into two extensions: one connected to dendrites that receive sensory information and another that transmits this information to the spinal cord.

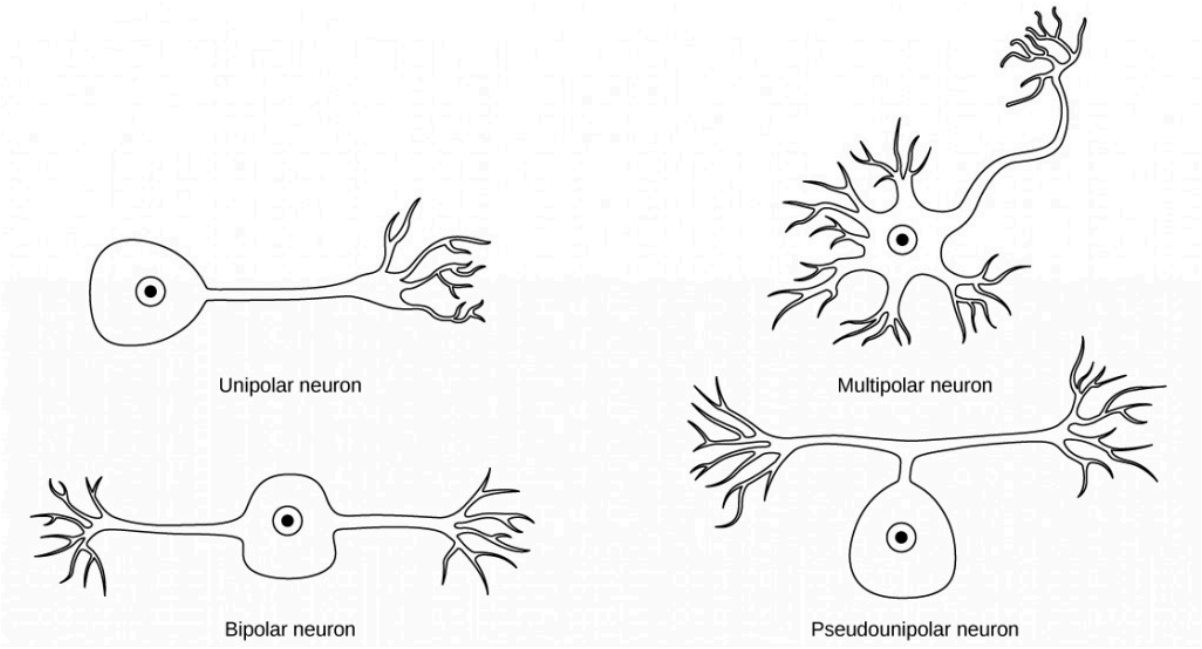


Figure 16.5. Neurons are broadly divided into four main types based on the number and placement of axons: (1) unipolar, (2) bipolar, (3) multipolar, and (4) pseudounipolar.

Neurogenesis

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

How do scientists identify new neurons? A researcher can inject a compound called bromodeoxyuridine (BrdU) into the brain of an animal. While all cells will be exposed to BrdU, BrdU will only be incorporated into the DNA of newly generated cells that are in S phase. A technique called immunohistochemistry can be used to attach a fluorescent label to the incorporated BrdU, and a researcher can use fluorescent microscopy to visualize the presence of BrdU, and thus new neurons, in brain tissue. Figure 16.6 is a micrograph which shows fluorescently labeled neurons in the hippocampus of a rat.

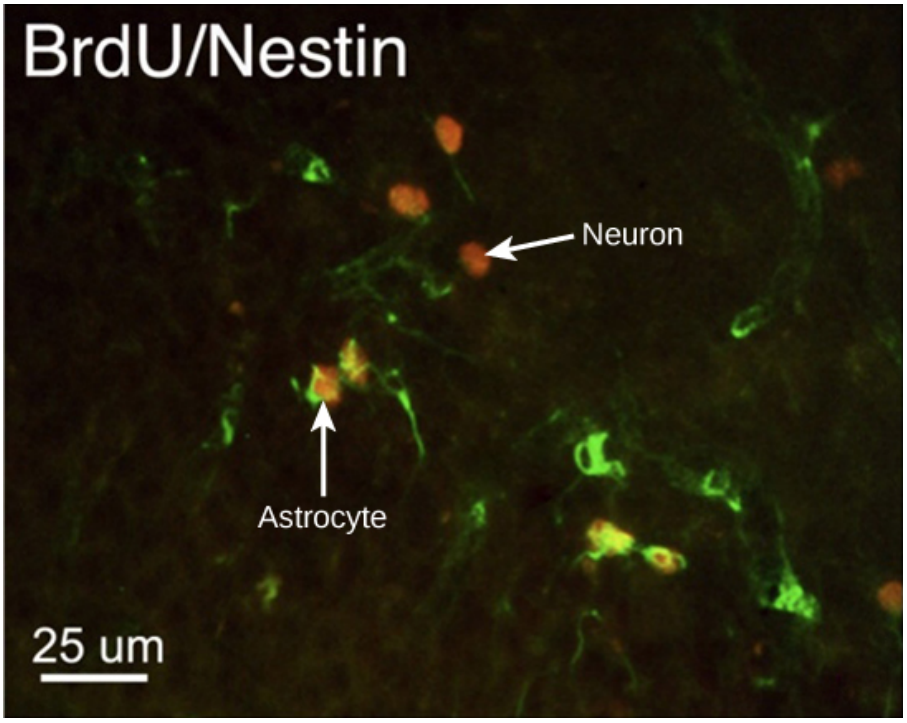


Figure 16.6. This micrograph shows fluorescently labeled new neurons in a rat hippocampus. Cells that are actively dividing have bromodoxyuridine (BrdU) incorporated into their DNA and are labeled in red. Cells that express glial fibrillary acidic protein (GFAP) are labeled in green. Astrocytes, but not neurons, express GFAP. Thus, cells that are labeled both red and green are actively dividing astrocytes, whereas cells labeled red only are actively dividing neurons. (credit: modification of work by Dr. Maryam Faiz, et. al., University of Barcelona; scale-bar data from Matt Russell)

Concept in Action

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

Glia

While glia are often thought of as the supporting cast of the nervous system, the number of glial cells in the brain actually outnumbers the number of neurons by a factor of ten. Neurons would be unable to function without the vital roles that are fulfilled by these glial cells. Glia guide developing neurons to their destinations, buffer ions and chemicals that would otherwise harm neurons, and provide myelin

sheaths around axons. Scientists have recently discovered that they also play a role in responding to nerve activity and modulating communication between nerve cells. When glia do not function properly, the result can be disastrous—most brain tumors are caused by mutations in glia.

Types of Glia

There are several different types of glia with different functions, two of which are shown in Figure 16.7. **Astrocytes**, shown in Figure 16.8**a** make contact with both capillaries and neurons in the CNS. They provide nutrients and other substances to neurons, regulate the concentrations of ions and chemicals in the extracellular fluid, and provide structural support for synapses. Astrocytes also form the blood-brain barrier—a structure that blocks entrance of toxic substances into the brain. Astrocytes, in particular, have been shown through calcium imaging experiments to become active in response to nerve activity, transmit calcium waves between astrocytes, and modulate the activity of surrounding synapses. **Satellite glia** provide nutrients and structural support for neurons in the PNS. **Microglia** scavenge and degrade dead cells and protect the brain from invading microorganisms. **Oligodendrocytes**, shown in Figure 16.8**b** form myelin sheaths around axons in the CNS. One axon can be myelinated by several oligodendrocytes, and one oligodendrocyte can provide myelin for multiple neurons. This is distinctive from the PNS where a single **Schwann cell** provides myelin for only one axon as the entire Schwann cell surrounds the axon. **Radial glia** serve as scaffolds for developing neurons as they migrate to their end destinations. **Ependymal** cells line fluid-filled ventricles of the brain and the central canal of the spinal cord. They are involved in the production of cerebrospinal fluid, which serves as a cushion for the brain, moves the fluid between the spinal cord and the brain, and is a component for the choroid plexus.

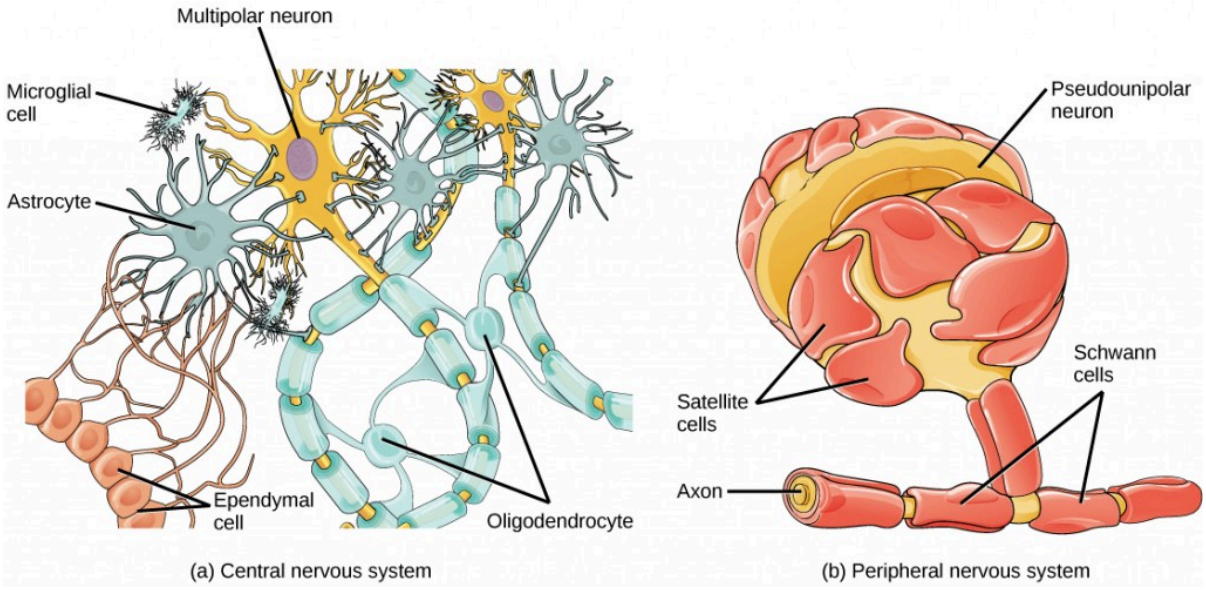


Figure 16.7.
Glial cells support neurons and maintain their environment. Glial cells of the (a) central nervous system include oligodendrocytes, astrocytes, ependymal cells, and microglial cells. Oligodendrocytes form the myelin sheath around axons. Astrocytes provide nutrients to neurons, maintain their extracellular environment, and provide structural support. Microglia scavenge pathogens and dead cells. Ependymal cells produce cerebrospinal fluid that cushions the neurons. Glial cells of the (b) peripheral nervous system include Schwann cells, which form the myelin sheath, and satellite cells, which provide nutrients and structural support to neurons.

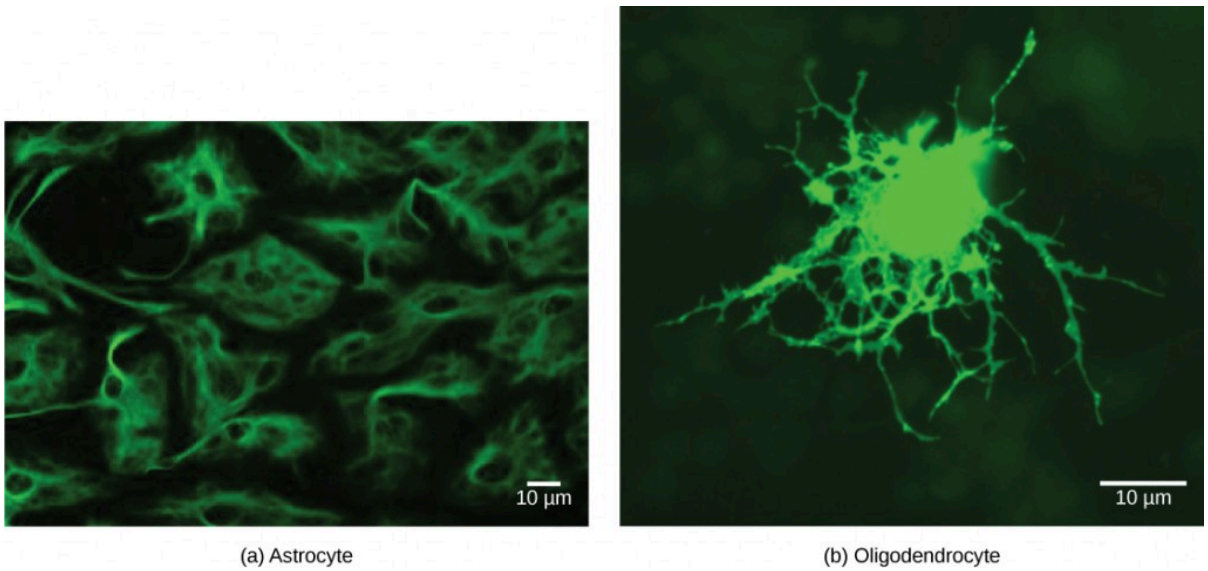


Figure 16.8.
(a) Astrocytes and (b) oligodendrocytes are glial cells of the central nervous system. (credit a: modification of work by Uniformed Services University; credit b: modification of work by Jurjen Broeke; scale-bar data from Matt Russell)

Summary

The nervous system is made up of neurons and glia. Neurons are specialized cells that are capable of sending electrical as well as chemical signals. Most neurons contain dendrites, which receive these signals, and axons that send signals to other neurons or tissues. There are four main types of neurons: unipolar, bipolar, multipolar, and pseudounipolar neurons. Glia are non-neuronal cells in the nervous system that support neuronal development and signaling. There are several types of glia that serve different functions.

Exercises

1. Which of the following statements is false?
    1. The soma is the cell body of a nerve cell.
    2. Myelin sheath provides an insulating layer to the dendrites.
    3. Axons carry the signal from the soma to the target.
    4. Dendrites carry the signal to the soma.
2. Neurons contain _____, which can receive signals from other neurons.
    1. axons
    2. mitochondria
    3. dendrites
    4. Golgi bodies
3. A(n) _____ neuron has one axon and one dendrite extending directly from the cell body.
    1. unipolar
    2. bipolar
    3. multipolar
    4. pseudounipolar
4. Glia that provide myelin for neurons in the brain are called _____.
    1. Schwann cells
    2. oligodendrocytes
    3. microglia
    4. astrocytes
5. How are neurons similar to other cells? How are they unique?
6. Multiple sclerosis causes demyelination of axons in the brain and spinal cord. Why is this problematic?

Answers

1. B
2. C
3. B
4. B
5. Neurons contain organelles common to all cells, such as a nucleus and mitochondria. They are unique because they contain dendrites, which can receive signals from other neurons, and axons that can send these signals to other cells.
6. Myelin provides insulation for signals traveling along axons. Without myelin, signal transmission can slow down and degrade over time. This would slow down neuronal communication across the

nervous system and affect all downstream functions.

Glossary

astrocyte: glial cell in the central nervous system that provide nutrients, extracellular buffering, and structural support for neurons; also makes up the blood-brain barrier
axon hillock: electrically sensitive structure on the cell body of a neuron that integrates signals from multiple neuronal connections
axon terminal: structure on the end of an axon that can form a synapse with another neuron
axon: tube-like structure that propagates a signal from a neuron's cell body to axon terminals
dendrite: structure that extends away from the cell body to receive messages from other neurons
ependymal: cell that lines fluid-filled ventricles of the brain and the central canal of the spinal cord; involved in production of
glia: (also, glial cells) cells that provide support functions for neurons
microglia: glia that scavenge and degrade dead cells and protect the brain from invading microorganisms
myelin: fatty substance produced by glia that insulates axons
neuron: specialized cell that can receive and transmit electrical and chemical signals
nodes of Ranvier: gaps in the myelin sheath where the signal is recharged
oligodendrocyte: glial cell that myelinates central nervous system neuron axons
radial glia: glia that serve as scaffolds for developing neurons as they migrate to their final destinations
Schwann cell: glial cell that creates myelin sheath around a peripheral nervous system neuron axon
satellite glia: glial cell that provides nutrients and structural support for neurons in the peripheral nervous system
synapse: junction between two neurons where neuronal signals are communicated

16.2 How Neurons Communicate

Learning Objectives

By the end of this section, you will be able to:

- Describe the basis of the resting membrane potential
- Explain the stages of an action potential and how action potentials are propagated
- Explain the similarities and differences between chemical and electrical synapses
- Describe long-term potentiation and long-term depression

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 16.9. Some ion channels need to be activated in order to open and allow ions to pass into or out of the cell. These ion channels are sensitive to the environment and can change their shape accordingly. Ion channels that change their structure in response to voltage changes are called voltage-gated ion channels. Voltage-gated ion channels regulate the relative concentrations of different ions inside and outside the cell. The difference in total charge between the inside and outside of the cell is called the **membrane potential**.

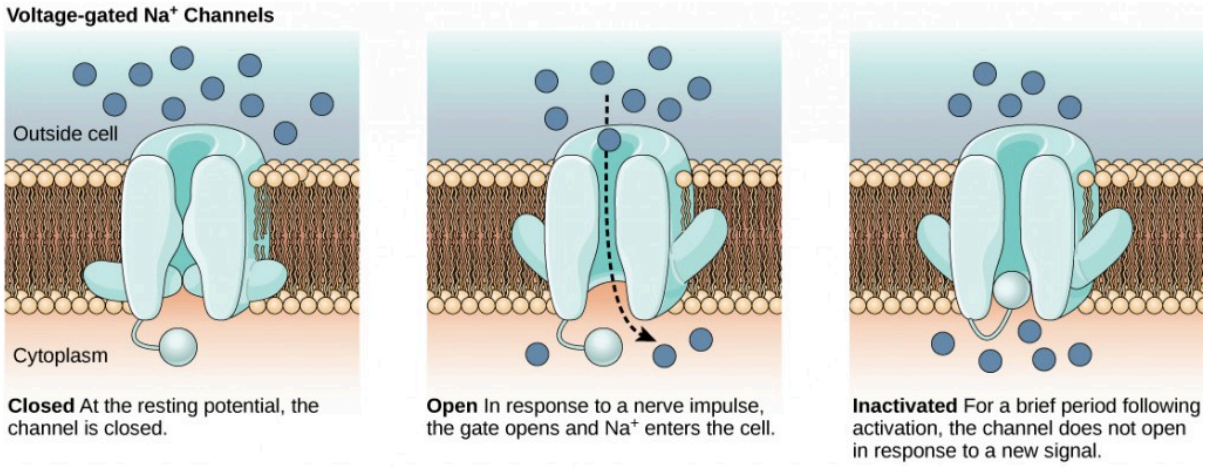


Figure 16.9. Voltage-gated ion channels open in response to changes in membrane voltage. After activation, they become inactivated for a brief period and will no longer open in response to a signal.

Concept in Action

This video discusses the basis of the resting membrane potential.

Resting Membrane Potential

A neuron at rest is negatively charged: the inside of a cell is approximately 70 millivolts more negative than the outside (−70 mV, note that this number varies by neuron type and by species). This voltage is called the resting membrane potential; it is caused by differences in the concentrations of ions inside and outside the cell. If the membrane were equally permeable to all ions, each type of ion would flow across the membrane and the system would reach equilibrium. Because ions cannot simply cross the membrane at will, there are different concentrations of several ions inside and outside the cell, as shown in Table 16.1. The difference in the number of positively charged potassium ions (K^+) inside and outside the cell dominates the resting membrane potential (Figure 16.10). When the membrane is at rest, K^+ ions accumulate inside the cell due to a net movement with the concentration gradient. The negative resting membrane potential is created and maintained by increasing the concentration of cations outside the cell (in the extracellular fluid) relative to inside the cell (in the cytoplasm). The negative charge within the cell is created by the cell membrane being more permeable to potassium ion movement than sodium ion movement. In neurons, potassium ions are maintained at high concentrations within the cell while sodium ions are maintained at high concentrations outside of the cell. The cell possesses potassium and sodium leakage channels that allow the two cations to diffuse down their concentration gradient. However, the neurons have far more potassium leakage channels

than sodium leakage channels. Therefore, potassium diffuses out of the cell at a much faster rate than sodium leaks in. Because more cations are leaving the cell than are entering, this causes the interior of the cell to be negatively charged relative to the outside of the cell. The actions of the sodium potassium pump help to maintain the resting potential, once established. Recall that sodium potassium pumps 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 calcium ions (Cl^-) tend to accumulate outside of the cell because they are repelled by negatively-charged proteins within the cytoplasm.

Table 16.1. The resting membrane potential is a result of different concentrations inside and outside the cell.

Ion Concentration Inside and Outside Neurons			
Ion	**Extracellular concentration (mM)**	**Intracellular concentration (mM)**	**Ratio outside/ inside**
Na^+	145	12	12
K+	4	155	0.026
Cl^-	120	4	30
Organic anions (A^-)	—	100	

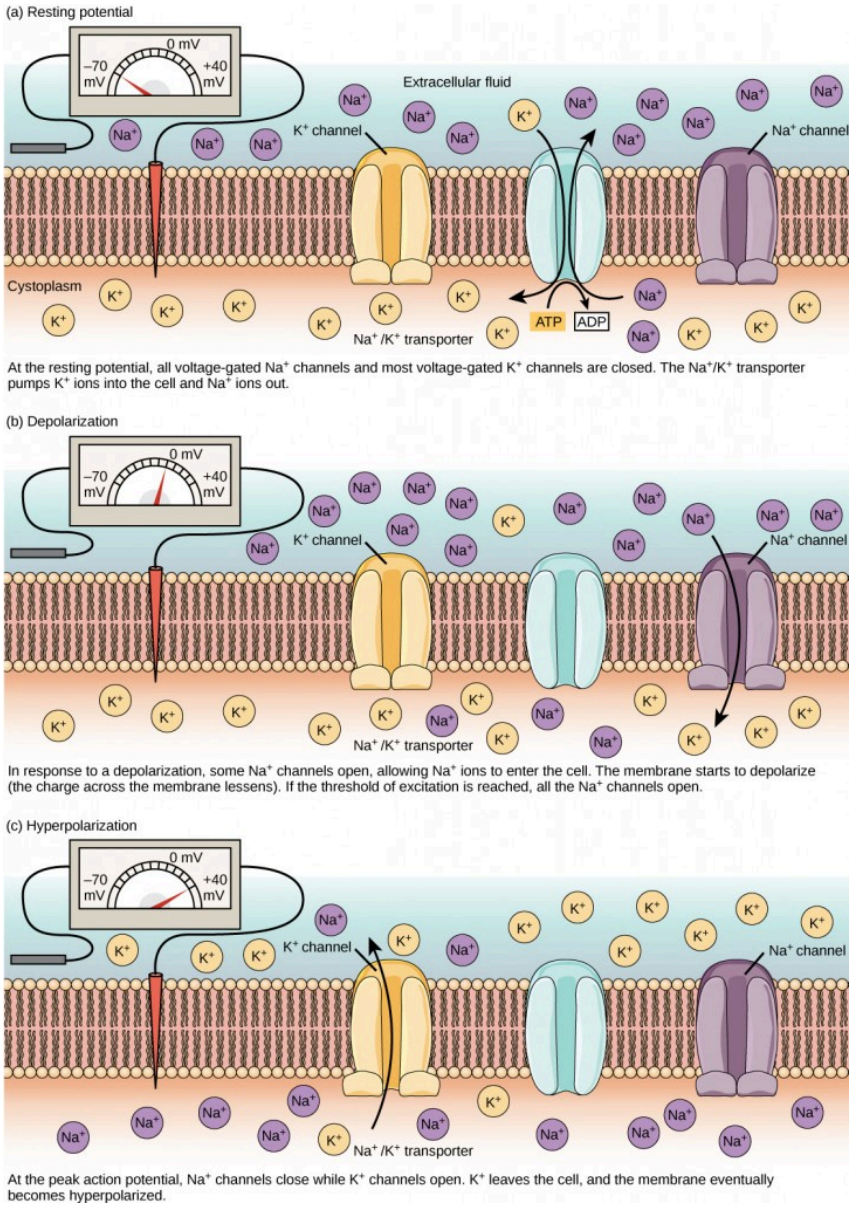


Figure 16.10.
The (a) resting membrane potential is a result of different concentrations of Na+ and K+ ions inside and outside the cell. A nerve impulse causes Na+ to enter the cell, resulting in (b) depolarization. At the peak action potential, K+ channels open and the cell becomes (c) hyperpolarized.

Action Potential

A neuron can receive input from other neurons and, if this input is strong enough, send the signal to downstream neurons. Transmission of a signal between neurons is generally carried by a chemical called a neurotransmitter. Transmission of a signal within a neuron (from dendrite to axon terminal) is carried by a brief reversal of the resting membrane potential called an **action potential**. When neurotransmitter molecules bind to receptors located on a neuron's dendrites, ion channels open. At excitatory synapses, this opening allows positive ions to enter the neuron and results in **depolarization** of the membrane—a decrease in the difference in voltage between the inside and outside of the neuron.

A stimulus from a sensory cell or another neuron depolarizes the target neuron to its threshold potential (-55 mV). Na^+ channels in the axon hillock open, allowing positive ions to enter the cell (Figure 16.10 and Figure 16.11). Once the sodium channels open, the neuron completely depolarizes to a membrane potential of about +40 mV. Action potentials are considered an "all-or nothing" event, in that, once the threshold potential is reached, the neuron always completely depolarizes. Once depolarization is complete, the cell must now "reset" its membrane voltage back to the resting potential. To accomplish this, the Na^+ channels close and cannot be opened. This begins the neuron's **refractory period**, in which it cannot produce another action potential because its sodium channels will not open. At the same time, voltage-gated K^+ channels open, allowing K^+ to leave the cell. As K^+ ions leave the cell, the membrane potential once again becomes negative. The diffusion of K^+ out of the cell actually **hyperpolarizes** the cell, in that the membrane potential becomes more negative than the cell's normal resting potential. At this point, the sodium channels will return to their resting state, meaning they are ready to open again if the membrane potential again exceeds the threshold potential. Eventually the extra K^+ ions diffuse out of the cell through the potassium leakage channels, bringing the cell from its hyperpolarized state, back to its resting membrane potential.

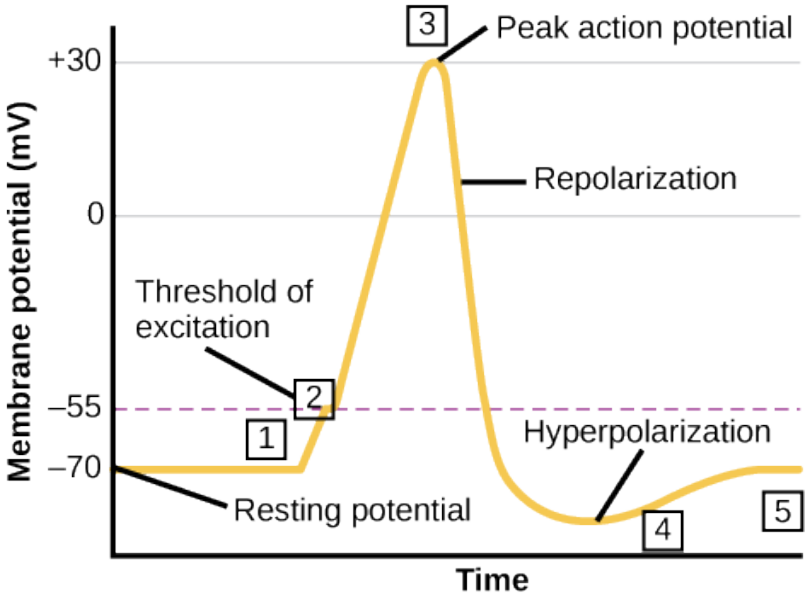


Figure 16.11. The formation of an action potential can be divided into five steps: (1) A stimulus from a sensory cell or another neuron causes the target cell to depolarize toward the threshold potential. (2) If the threshold of excitation is reached, all Na+ channels open and the membrane depolarizes. (3) At the peak action potential, K+ channels open and K+ begins to leave the cell. At the same time, Na+ channels close. (4) The membrane becomes hyperpolarized as K+ ions continue to leave the cell. The hyperpolarized membrane is in a refractory period and cannot fire. (5) The K+ channels close and the Na+/K+ transporter restores the resting potential.

Potassium channel blockers, such as amiodarone and procainamide, which are used to treat abnormal electrical activity in the heart, called cardiac dysrhythmia, impede the movement of K^+ through voltage-gated K^+ channels. Which part of the action potential would you expect potassium channels to affect?

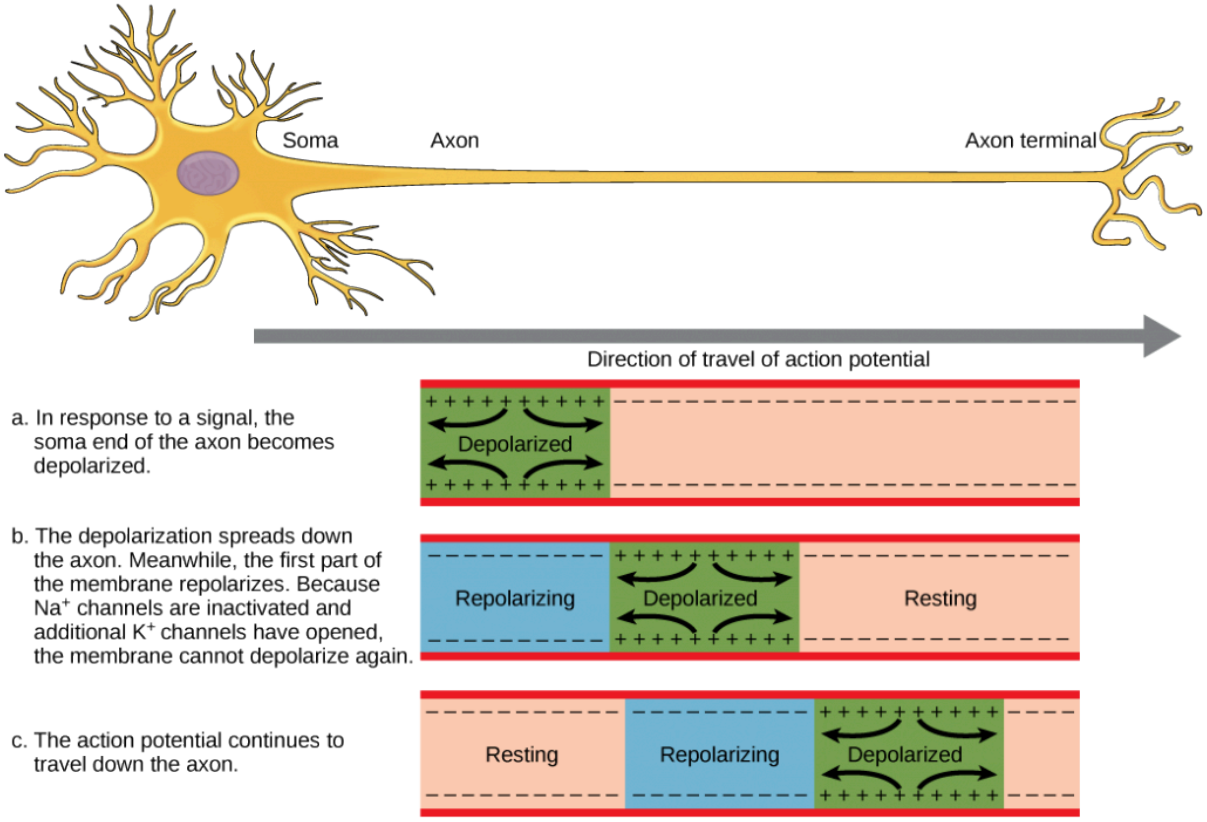


Figure 16.12. The action potential is conducted down the axon as the axon membrane depolarizes, then repolarizes.

Concept in Action

This video presents an overview of action potential.

Myelin and the Propagation of the Action Potential

For an action potential to communicate information to another neuron, it must travel along the axon and reach the axon terminals where it can initiate neurotransmitter release. The speed of conduction of an action potential along an axon is influenced by both the diameter of the axon and the axon's resistance to current leak. Myelin acts as an insulator that prevents current from leaving the axon; this increases the speed of action potential conduction. In demyelinating diseases like multiple sclerosis, action potential conduction slows because current leaks from previously insulated axon areas. The nodes of Ranvier, illustrated in Figure 16.13 are gaps in the myelin sheath along the axon. These

unmyelinated spaces are about one micrometer long and contain voltage gated Na^+ and K^+ channels. Flow of ions through these channels, particularly the Na^+ channels, regenerates the action potential over and over again along the axon. This 'jumping' of the action potential from one node to the next is called **saltatory conduction**. If nodes of Ranvier were not present along an axon, the action potential would propagate very slowly since Na^+ and K^+ channels would have to continuously regenerate action potentials at every point along the axon instead of at specific points. Nodes of Ranvier also save energy for the neuron since the channels only need to be present at the nodes and not along the entire axon.

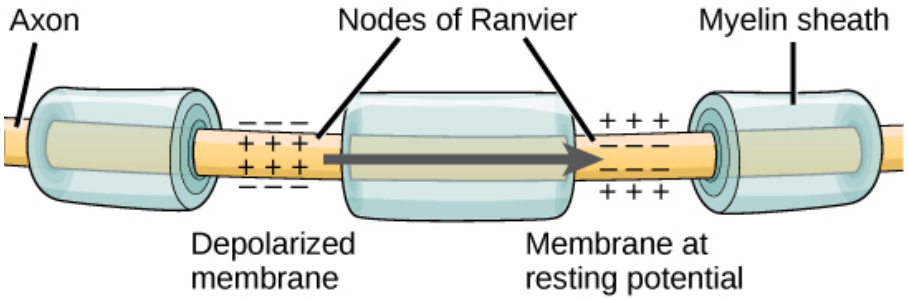


Figure 16.13. Nodes of Ranvier are gaps in myelin coverage along axons. Nodes contain voltage-gated K+ and Na+ channels. Action potentials travel down the axon by jumping from one node to the next.

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 16.14, which is an image from a scanning electron microscope.

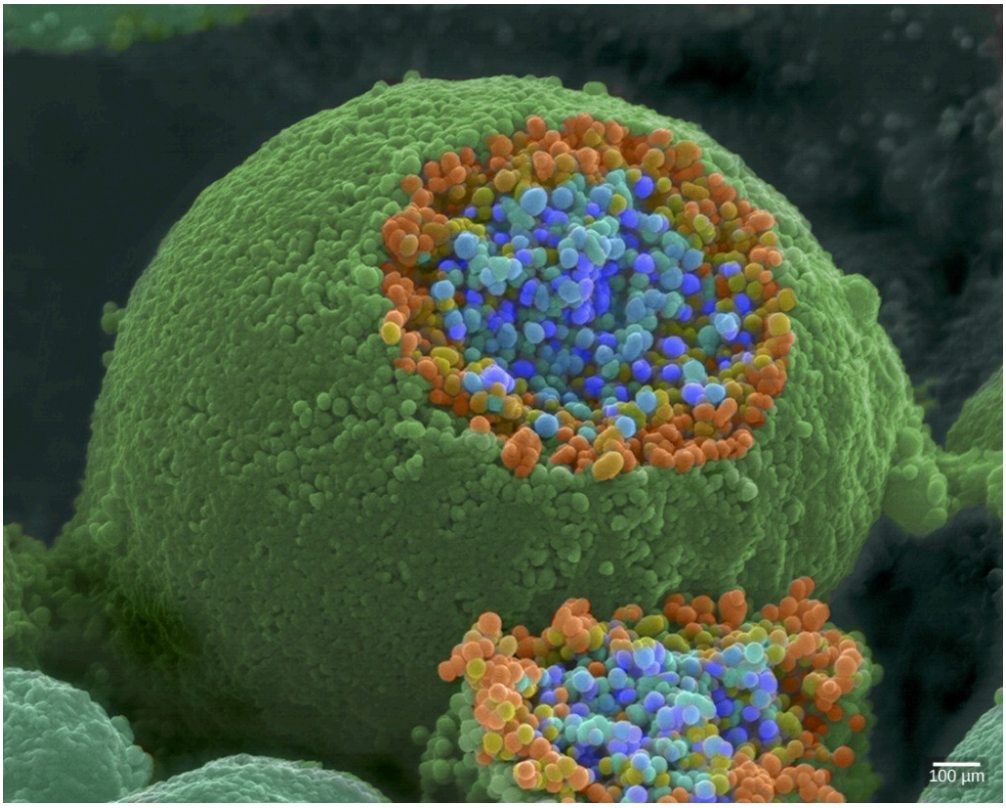


Figure 16.14. This pseudocolored image taken with a scanning electron microscope shows an axon terminal that was broken open to reveal synaptic vesicles (blue and orange) inside the neuron. (credit: modification of work by Tina Carvalho, NIH-NIGMS; scale-bar data from Matt Russell)

Fusion of a vesicle with the presynaptic membrane causes neurotransmitter to be released into the **synaptic cleft**, the extracellular space between the presynaptic and postsynaptic membranes, as illustrated in Figure 16.15. The neurotransmitter diffuses across the synaptic cleft and binds to receptor proteins on the postsynaptic membrane.

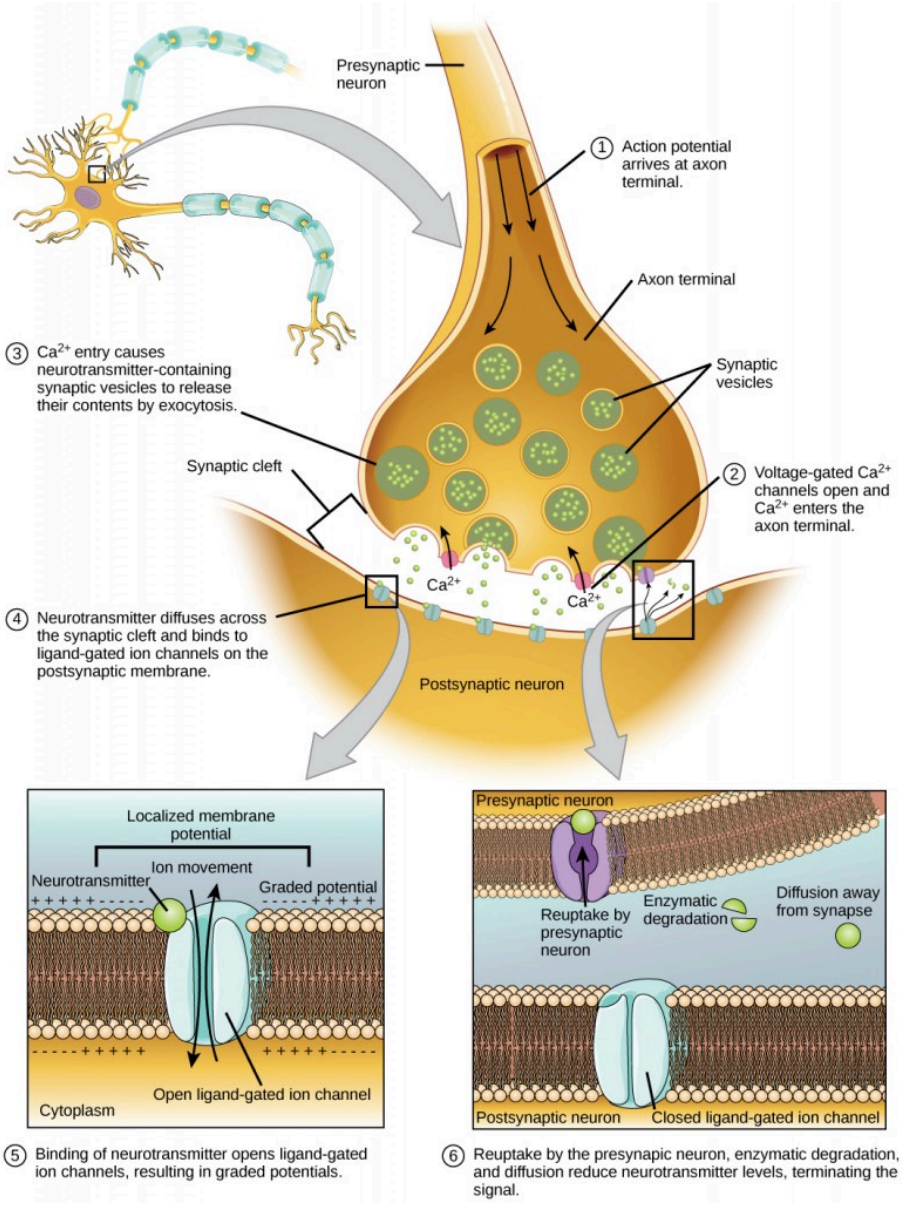


Figure 16.15. Communication at chemical synapses requires release of neurotransmitters. When the presynaptic membrane is depolarized, voltage-gated Ca2+ channels open and allow Ca2+ to enter the cell. The calcium entry causes synaptic vesicles to fuse with the membrane and release neurotransmitter molecules into the synaptic cleft. The neurotransmitter diffuses across the synaptic cleft and binds to ligand-gated ion channels in the postsynaptic membrane, resulting in a localized depolarization or hyperpolarization of the postsynaptic neuron.

The binding of a specific neurotransmitter causes particular ion channels, in this case ligand-gated channels, on the postsynaptic membrane to open. Neurotransmitters can either have excitatory or inhibitory effects on the postsynaptic membrane, as detailed in Table 16.2. For example, when acetylcholine is released at the synapse between a nerve and muscle (called the neuromuscular junction) by a presynaptic neuron, it causes postsynaptic Na^+ channels to open. Na^+ enters the postsynaptic cell and causes the postsynaptic membrane to depolarize. This depolarization is called an **excitatory postsynaptic potential (EPSP)** and makes the postsynaptic neuron more likely to fire an action potential. Release of neurotransmitter at inhibitory synapses causes **inhibitory postsynaptic**

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

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

Table 16.2.Neurotransmitter Function and Location

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

Electrical Synapse

While electrical synapses are fewer in number than chemical synapses, they are found in all nervous systems and play important and unique roles. The mode of neurotransmission in electrical synapses is quite different from that in chemical synapses. In an electrical synapse, the presynaptic and postsynaptic membranes are very close together and are actually physically connected by channel proteins forming gap junctions. Gap junctions allow current to pass directly from one cell to the next. In addition to the ions that carry this current, other molecules, such as ATP, can diffuse through the large gap junction pores.

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

Signal Summation

Sometimes a single EPSP is strong enough to induce an action potential in the postsynaptic neuron, but often multiple presynaptic inputs must create EPSPs around the same time for the postsynaptic neuron to be sufficiently depolarized to fire an action potential. This process is called **summation** and occurs at the axon hillock, as illustrated in Figure 16.16. Additionally, one neuron often has inputs from many presynaptic neurons—some excitatory and some inhibitory—so IPSPs can cancel out EPSPs and vice versa. It is the net change in postsynaptic membrane voltage that determines whether the postsynaptic cell has reached its threshold of excitation needed to fire an action potential. Together, synaptic summation and the threshold for excitation act as a filter so that random "noise" in the system is not transmitted as important information.

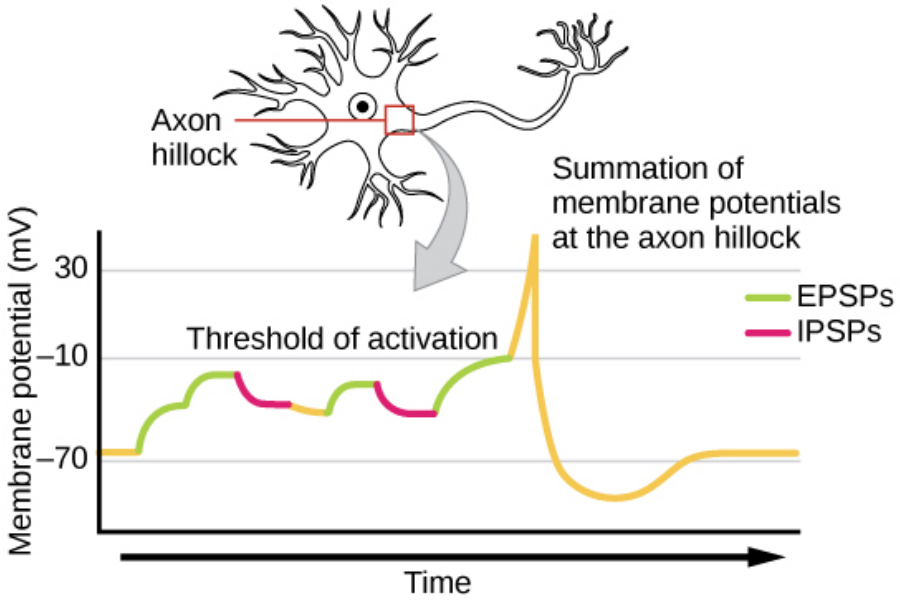


Figure 16.16. A single neuron can receive both excitatory and inhibitory inputs from multiple neurons, resulting in local membrane depolarization (EPSP input) and hyperpolarization (IPSP input). All these inputs are added together at the axon hillock. If the EPSPs are strong enough to overcome the IPSPs and reach the threshold of excitation, the neuron will fire.

Brain-computer interface

Amyotrophic lateral sclerosis (ALS, also called Lou Gehrig's Disease) is a neurological disease characterized by the degeneration of the motor neurons that control voluntary movements. The disease begins with muscle weakening and lack of coordination and eventually destroys the neurons that control speech, breathing, and swallowing; in the end, the disease can lead to paralysis. At that point, patients require assistance from machines to be able to breathe and to communicate. Several special technologies have been developed to allow "locked-in" patients to communicate with the rest of the world. One technology, for example, allows patients to type out sentences by twitching their cheek. These sentences can then be read aloud by a computer.

A relatively new line of research for helping paralyzed patients, including those with ALS, to communicate and retain a degree of self-sufficiency is called brain-computer interface (BCI)

technology and is illustrated in Figure 16.17. This technology sounds like something out of science fiction: it allows paralyzed patients to control a computer using only their thoughts. There are several forms of BCI. Some forms use EEG recordings from electrodes taped onto the skull. These recordings contain information from large populations of neurons that can be decoded by a computer. Other forms of BCI require the implantation of an array of electrodes smaller than a postage stamp in the arm and hand area of the motor cortex. This form of BCI, while more invasive, is very powerful as each electrode can record actual action potentials from one or more neurons. These signals are then sent to a computer, which has been trained to decode the signal and feed it to a tool—such as a cursor on a computer screen. This means that a patient with ALS can use e-mail, read the Internet, and communicate with others by thinking of moving his or her hand or arm (even though the paralyzed patient cannot make that bodily movement). Recent advances have allowed a paralyzed locked-in patient who suffered a stroke 15 years ago to control a robotic arm and even to feed herself coffee using BCI technology.

Despite the amazing advancements in BCI technology, it also has limitations. The technology can require many hours of training and long periods of intense concentration for the patient; it can also require brain surgery to implant the devices.

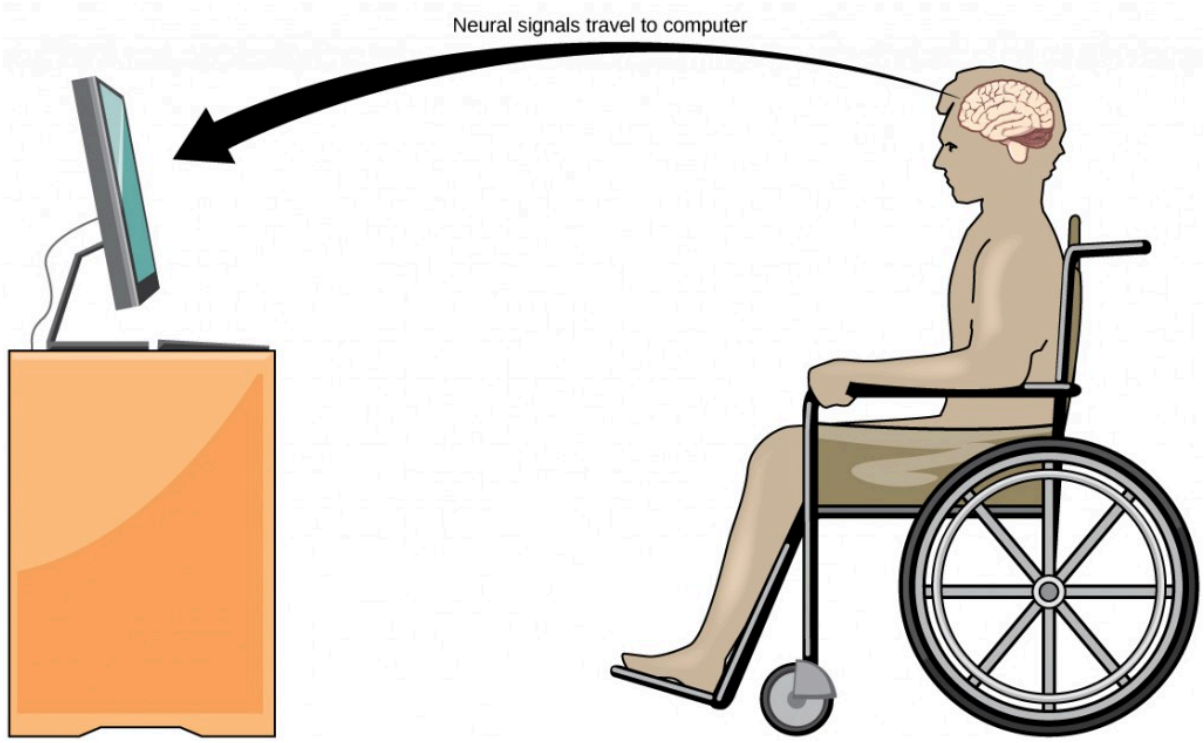


Figure 16.17. With brain-computer interface technology, neural signals from a paralyzed patient are collected, decoded, and then fed to a tool, such as a computer, a wheelchair, or a robotic arm.

Concept in Action

Watch this video in which a paralyzed woman use a brain-controlled robotic arm to bring a drink to her mouth, among other images of brain-computer interface technology in action.

Synaptic Plasticity

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

Long-term Potentiation (LTP)

Long-term potentiation (LTP) is a persistent strengthening of a synaptic connection. LTP is based on the Hebbian principle: cells that fire together wire together. There are various mechanisms, none fully understood, behind the synaptic strengthening seen with LTP. One known mechanism involves a type of postsynaptic glutamate receptor, called NMDA (N-Methyl-D-aspartate) receptors, shown in Figure 16.18. These receptors are normally blocked by magnesium ions; however, when the postsynaptic neuron is depolarized by multiple presynaptic inputs in quick succession (either from one neuron or multiple neurons), the magnesium ions are forced out allowing Ca ions to pass into the postsynaptic cell. Next, Ca^{2+} ions entering the cell initiate a signaling cascade that causes a different type of glutamate receptor, called AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) receptors, to be inserted into the postsynaptic membrane, since activated AMPA receptors allow positive ions to enter the cell. So, the next time glutamate is released from the presynaptic membrane, it will have a larger excitatory effect (EPSP) on the postsynaptic cell because the binding of glutamate to these AMPA receptors will allow more positive ions into the cell. The insertion of additional AMPA receptors strengthens the synapse and means that the postsynaptic neuron is more likely to fire in response to presynaptic neurotransmitter release. Some drugs of abuse co-opt the LTP pathway, and this synaptic strengthening can lead to addiction.

Long-term Depression (LTD)

Long-term depression (LTD) is essentially the reverse of LTP: it is a long-term weakening of a synaptic connection. One mechanism known to cause LTD also involves AMPA receptors. In this situation, calcium that enters through NMDA receptors initiates a different signaling cascade, which

results in the removal of AMPA receptors from the postsynaptic membrane, as illustrated in Figure 16.18. The decrease in AMPA receptors in the membrane makes the postsynaptic neuron less responsive to glutamate released from the presynaptic neuron. While it may seem counterintuitive, LTD may be just as important for learning and memory as LTP. The weakening and pruning of unused synapses allows for unimportant connections to be lost and makes the synapses that have undergone LTP that much stronger by comparison.

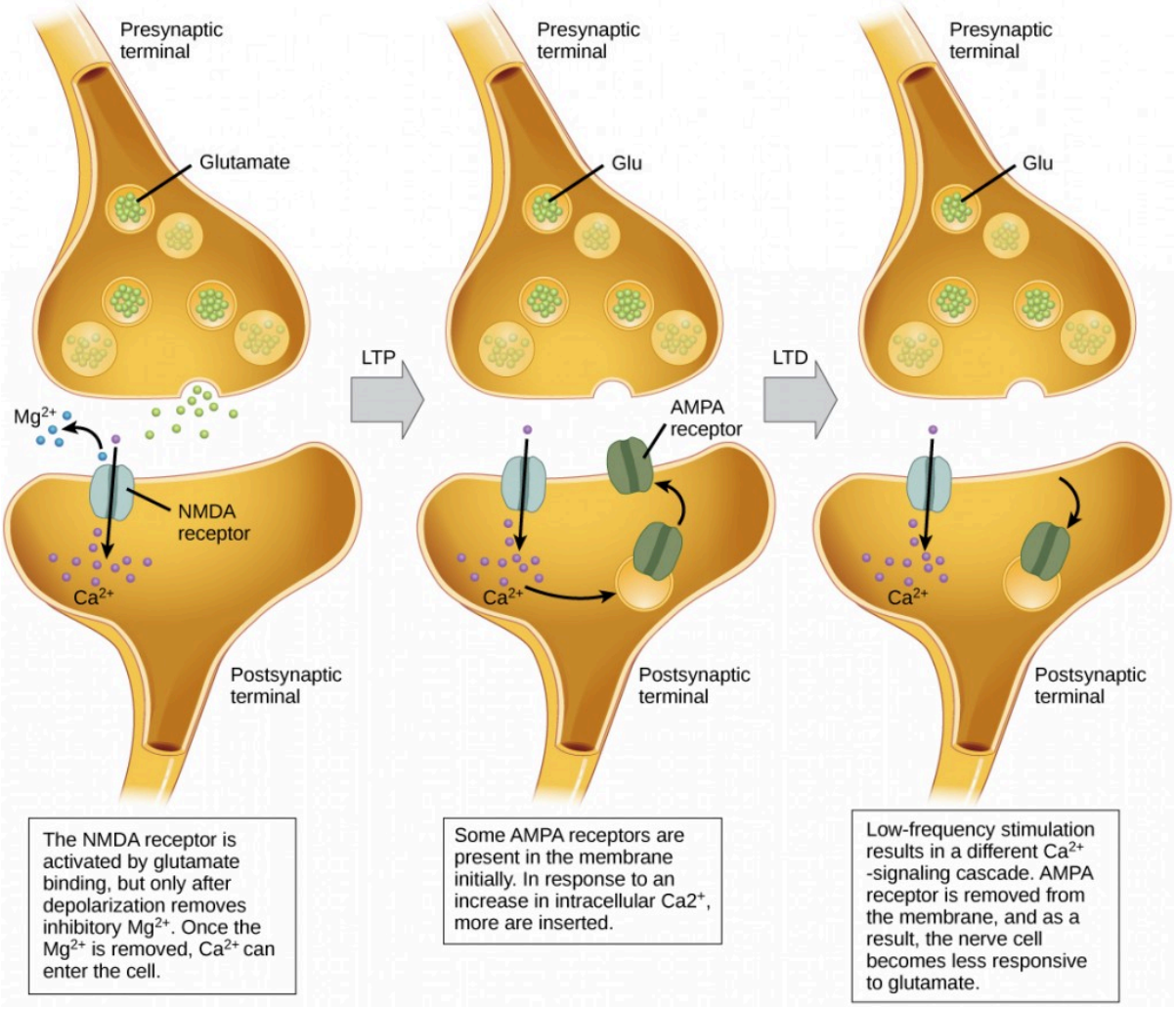


Figure 16.18. Calcium entry through postsynaptic NMDA receptors can initiate two different forms of synaptic plasticity: long-term potentiation (LTP) and long-term depression (LTD). LTP arises when a single synapse is repeatedly stimulated. This stimulation causes a calcium- and CaMKII-dependent cellular cascade, which results in the insertion of more AMPA receptors into the postsynaptic membrane. The next time glutamate is released from the presynaptic cell, it will bind to both NMDA and the newly inserted AMPA receptors, thus depolarizing the membrane more efficiently. LTD occurs when few glutamate molecules bind to NMDA receptors at a synapse (due to a low firing rate of the presynaptic neuron). The calcium that does flow through NMDA receptors initiates a different calcineurin and protein phosphatase 1-dependent cascade, which results in the endocytosis of AMPA receptors. This makes the postsynaptic neuron less responsive to glutamate released from the presynaptic neuron.

Summary

Neurons have charged membranes because there are different concentrations of ions inside and outside

of the cell. Voltage-gated ion channels control the movement of ions into and out of a neuron. When a neuronal membrane is depolarized to at least the threshold of excitation, an action potential is fired. The action potential is then propagated along a myelinated axon to the axon terminals. In a chemical synapse, the action potential causes release of neurotransmitter molecules into the synaptic cleft. Through binding to postsynaptic receptors, the neurotransmitter can cause excitatory or inhibitory postsynaptic potentials by depolarizing or hyperpolarizing, respectively, the postsynaptic membrane. In electrical synapses, the action potential is directly communicated to the postsynaptic cell through gap junctions—large channel proteins that connect the pre-and postsynaptic membranes. Synapses are not static structures and can be strengthened and weakened. Two mechanisms of synaptic plasticity are long-term potentiation and long-term depression.

Exercises

1. Potassium channel blockers, such as amiodarone and procainamide, which are used to treat abnormal electrical activity in the heart, called cardiac dysrhythmia, impede the movement of K+ through voltage-gated K+ channels. Which part of the action potential would you expect potassium channels to affect?
2. For a neuron to fire an action potential, its membrane must reach _____.
    1. hyperpolarization
    2. the threshold of excitation
    3. the refractory period
    4. inhibitory postsynaptic potential
3. After an action potential, the opening of additional voltage-gated _____ channels and the inactivation of sodium channels, cause the membrane to return to its resting membrane potential.
    1. sodium
    2. potassium
    3. calcium
    4. chloride
4. What is the term for protein channels that connect two neurons at an electrical synapse?
    1. synaptic vesicles
    2. voltage-gated ion channels
    3. gap junction protein
    4. sodium-potassium exchange pumps
5. How does myelin aid propagation of an action potential along an axon? How do the nodes of Ranvier help this process?
6. What are the main steps in chemical neurotransmission?

Answers

1. Potassium channel blockers slow the repolarization phase, but have no effect on depolarization.

2. B
3. B
4. C
5. Myelin prevents the leak of current from the axon. Nodes of Ranvier allow the action potential to be regenerated at specific points along the axon. They also save energy for the cell since voltage-gated ion channels and sodium-potassium transporters are not needed along myelinated portions of the axon.
6. An action potential travels along an axon until it depolarizes the membrane at an axon terminal. Depolarization of the membrane causes voltage-gated Ca^{2+} channels to open and Ca^{2+} to enter the cell. The intracellular calcium influx causes synaptic vesicles containing neurotransmitter to fuse with the presynaptic membrane. The neurotransmitter diffuses across the synaptic cleft and binds to receptors on the postsynaptic membrane. Depending on the specific neurotransmitter and postsynaptic receptor, this action can cause positive (excitatory postsynaptic potential) or negative (inhibitory postsynaptic potential) ions to enter the cell.

Glossary

action potential
self-propagating momentary change in the electrical potential of a neuron (or muscle) membrane

depolarization
change in the membrane potential to a less negative value

excitatory postsynaptic potential (EPSP)
depolarization of a postsynaptic membrane caused by neurotransmitter molecules released from a presynaptic cell

hyperpolarization
change in the membrane potential to a more negative value

inhibitory postsynaptic potential (IPSP)
hyperpolarization of a postsynaptic membrane caused by neurotransmitter molecules released from a presynaptic cell

long-term depression (LTD)
prolonged decrease in synaptic coupling between a pre- and postsynaptic cell

membrane potential
difference in electrical potential between the inside and outside of a cell

refractory period
period after an action potential when it is more difficult or impossible for an action potential to be fired; caused by inactivation of sodium channels and activation of additional potassium channels of the membrane

saltatory conduction
"jumping" of an action potential along an axon from one node of Ranvier to the next

summation
process of multiple presynaptic inputs creating EPSPs around the same time for the postsynaptic neuron to be sufficiently depolarized to fire an action potential

synaptic cleft
space between the presynaptic and postsynaptic membranes

synaptic vesicle
spherical structure that contains a neurotransmitter

16.3 The Central Nervous System

Learning Objectives

By the end of this section, you will be able to:

- Identify the spinal cord, cerebral lobes, and other brain areas on a diagram of the brain
- Describe the basic functions of the spinal cord, cerebral lobes, and other brain areas

The central nervous system (CNS) is made up of the brain, a part of which is shown in Figure 16.19 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.

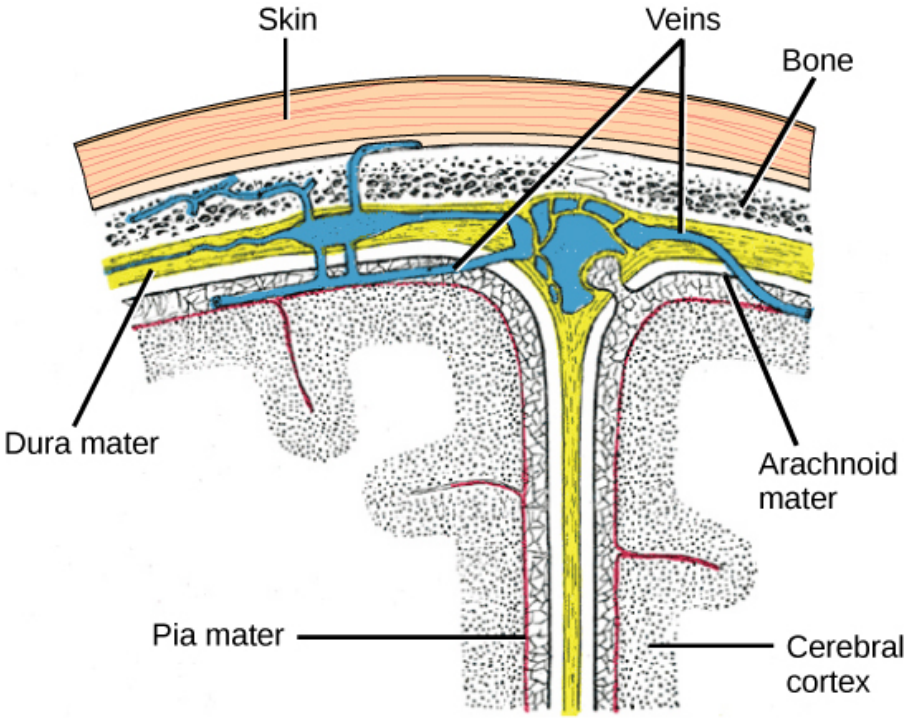


Figure 16.19. The cerebral cortex is covered by three layers of meninges: the dura, arachnoid, and pia maters. (credit: modification of work by Gray's Anatomy)

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 16.21**b**, a coronal section cuts the brain front to back, as shown in Figure 16.20**a**, 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.

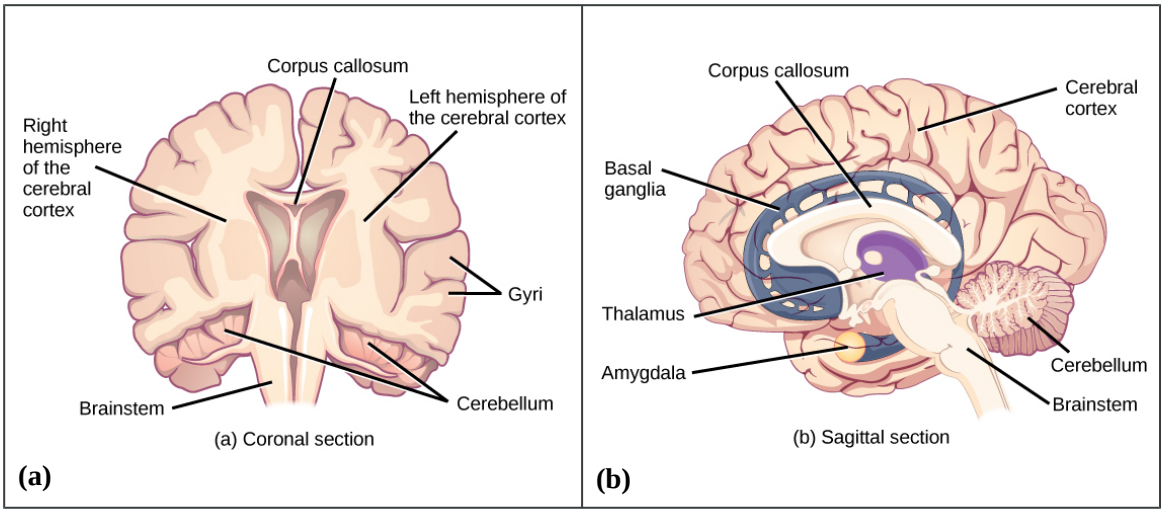


Figure 16.20 These illustrations show the (a) coronal and (b) sagittal sections of the human brain.

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

Concept in Action

See this website to learn more about split-brain patients and to play a game where you can model the split-brain experiments yourself.

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 16.21 illustrates these four lobes of the human cerebral cortex.

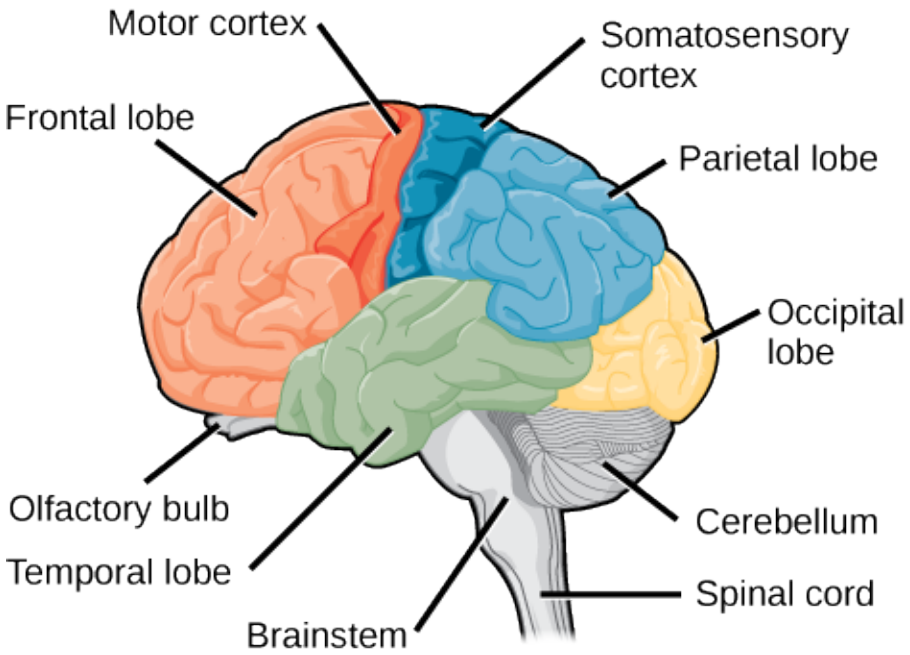


Figure 16.21. The human cerebral cortex includes the frontal, parietal, temporal, and occipital lobes.

The **frontal lobe** is located at the front of the brain, over the eyes. This lobe contains the olfactory bulb, which processes smells. The frontal lobe also contains the motor cortex, which is important for planning and implementing movement. Areas within the motor cortex map to different muscle groups, and there is some organization to this map, as shown in Figure 16.22. 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.

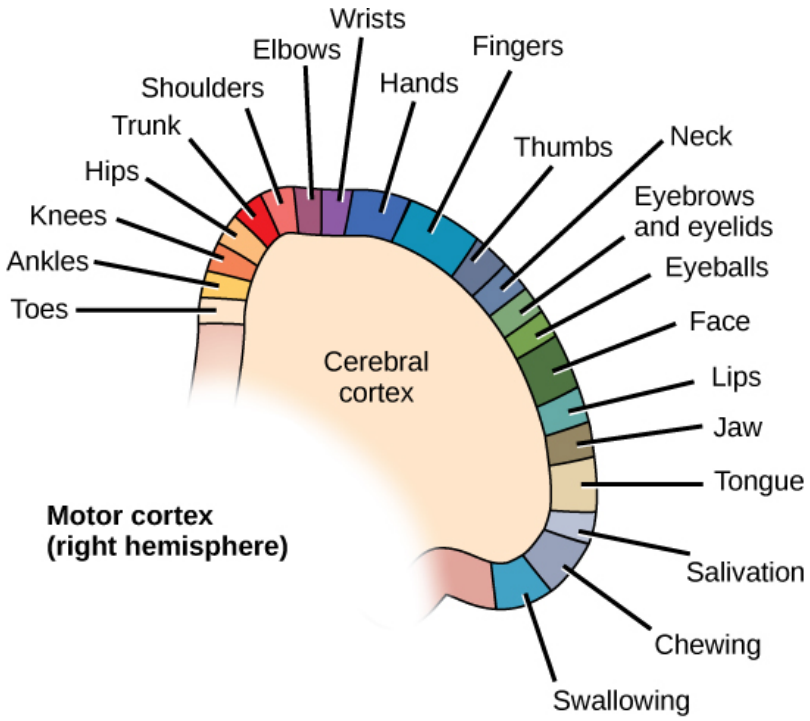


Figure 16.22. Different parts of the motor cortex control different muscle groups. Muscle groups that are neighbors in the body are generally controlled by neighboring regions of the motor cortex as well. For example, the neurons that control finger movement are near the neurons that control hand movement.

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

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

The **temporal lobe** is located at the base of the brain by your ears and is primarily involved in processing and interpreting sounds. It also contains the **hippocampus** (Greek for "seahorse")—a structure that processes memory formation. The hippocampus is illustrated in Figure 16.24. 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).

Cerebral Cortex

Compared to other vertebrates, mammals have exceptionally large brains for their body size. An entire alligator's brain, for example, would fill about one and a half teaspoons. This increase in brain to body size ratio is especially pronounced in apes, whales, and dolphins. While this increase in overall brain

size doubtlessly played a role in the evolution of complex behaviors unique to mammals, it does not tell the whole story. Scientists have found a relationship between the relatively high surface area of the cortex and the intelligence and complex social behaviors exhibited by some mammals. This increased surface area is due, in part, to increased folding of the cortical sheet (more sulci and gyri). For example, a rat cortex is very smooth with very few sulci and gyri. Cat and sheep cortices have more sulci and gyri. Chimps, humans, and dolphins have even more.

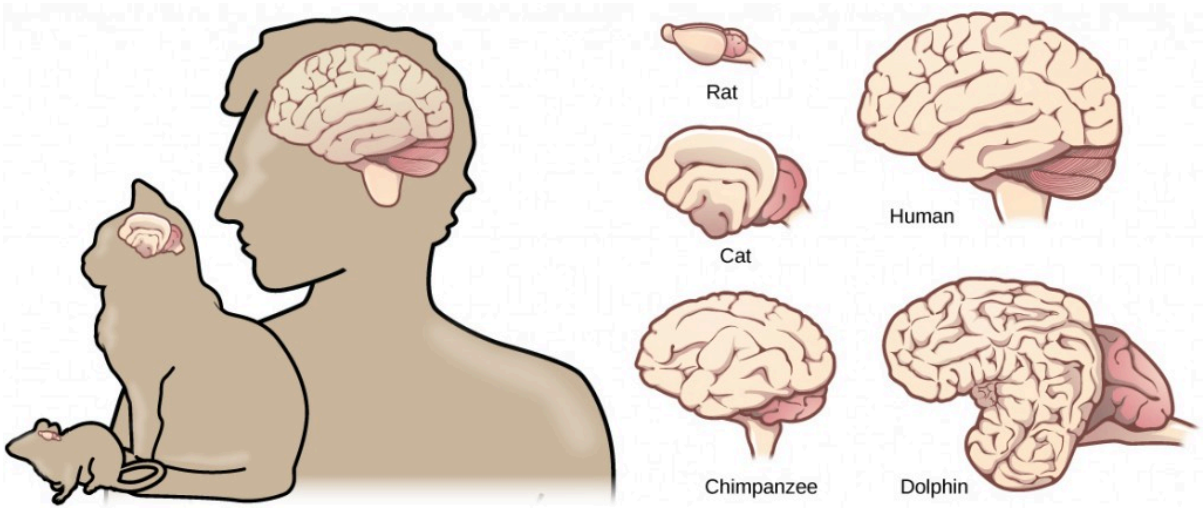


Figure 16.23. Mammals have larger brain-to-body ratios than other vertebrates. Within mammals, increased cortical folding and surface area is correlated with complex behavior.

Basal Ganglia

Interconnected brain areas called the **basal ganglia (or basal nuclei),** shown in Figure 16.20**b**, play important roles in movement control and posture. Damage to the basal ganglia, as in Parkinson's disease, leads to motor impairments like a shuffling gait when walking. The basal ganglia also regulate motivation. For example, when a wasp sting led to bilateral basal ganglia damage in a 25-year-old businessman, he began to spend all his days in bed and showed no interest in anything or anybody. But when he was externally stimulated—as when someone asked to play a card game with him—he was able to function normally. Interestingly, he and other similar patients do not report feeling bored or frustrated by their state.

Thalamus

The **thalamus** (Greek for "inner chamber"), illustrated in Figure 16.24, acts as a gateway to and from the cortex. It receives sensory and motor inputs from the body and also receives feedback from the cortex. This feedback mechanism can modulate conscious awareness of sensory and motor inputs depending on the attention and arousal state of the animal. The thalamus helps regulate consciousness, arousal, and sleep states. A rare genetic disorder called fatal familial insomnia causes the degeneration of thalamic neurons and glia. This disorder prevents affected patients from being able to sleep, among other symptoms, and is eventually fatal.

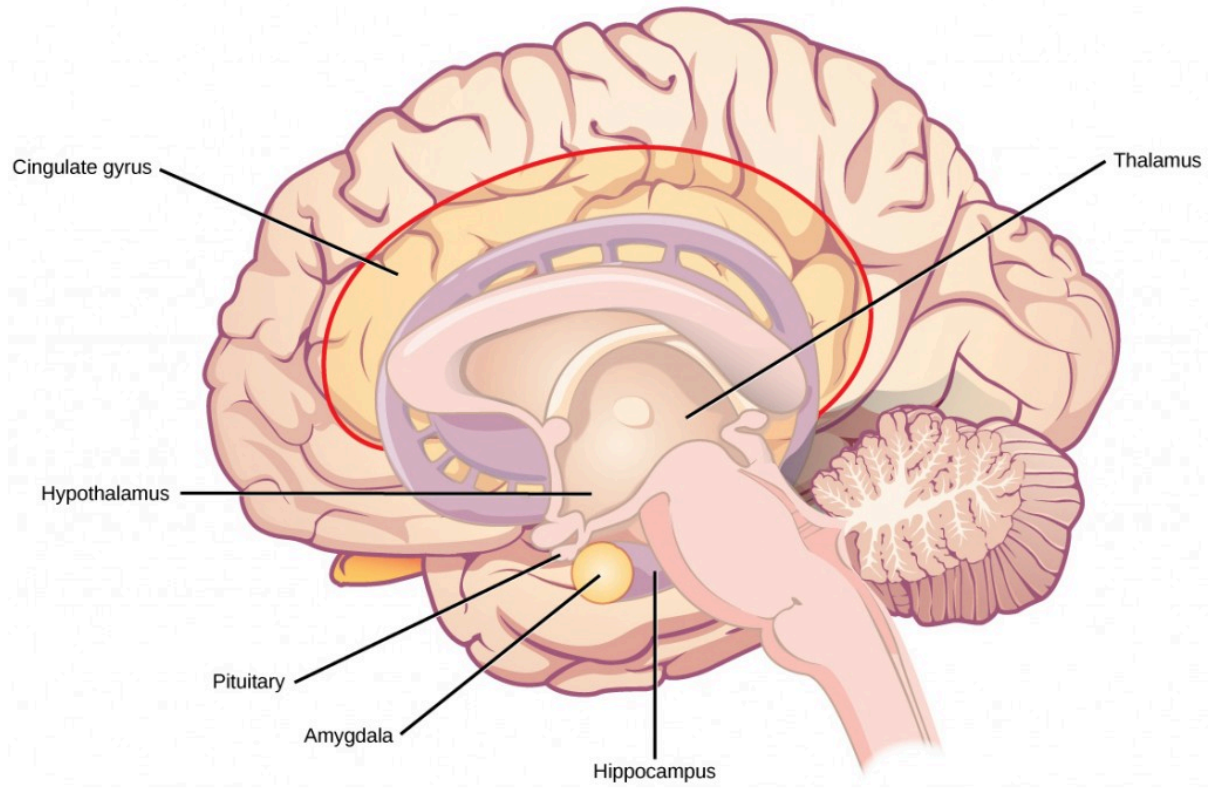


Figure 16.24. The limbic system regulates emotion and other behaviors. It includes parts of the cerebral cortex located near the center of the brain, including the cingulate gyrus and the hippocampus as well as the thalamus, hypothalamus and amygdala.

Hypothalamus

Below the thalamus is the **hypothalamus**, shown in Figure 16.24. The hypothalamus controls the endocrine system by sending signals to the pituitary gland, a pea-sized endocrine gland that releases several different hormones that affect other glands as well as other cells. This relationship means that the hypothalamus regulates important behaviors that are controlled by these hormones. The hypothalamus is the body's thermostat—it makes sure key functions like food and water intake, energy expenditure, and body temperature are kept at appropriate levels. Neurons within the hypothalamus also regulate circadian rhythms, sometimes called sleep cycles.

Limbic System

The **limbic system** is a connected set of structures that regulates emotion, as well as behaviors related to fear and motivation. It plays a role in memory formation and includes parts of the thalamus and hypothalamus as well as the hippocampus. One important structure within the limbic system is a temporal lobe structure called the **amygdala** (Greek for "almond"), illustrated in Figure 16.24. The two amygdala are important both for the sensation of fear and for recognizing fearful faces. The **cingulate gyrus** helps regulate emotions and pain.

Cerebellum

The **cerebellum** (Latin for "little brain"), shown in Figure 16.21, 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 16.21, 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 16.21. 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 16.25. 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.

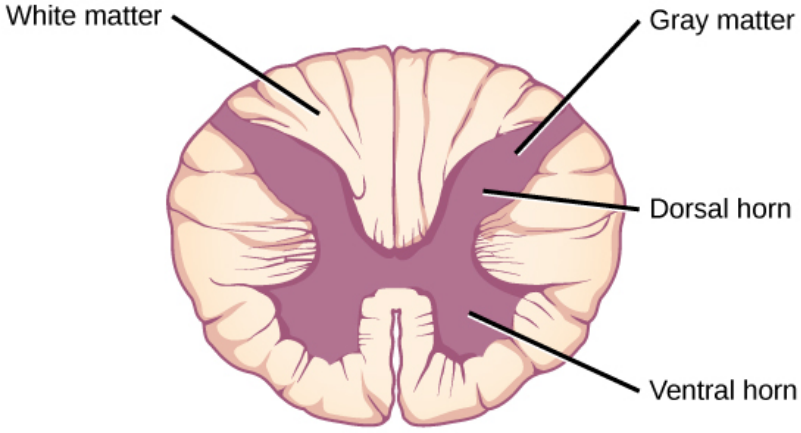


Figure 16.25. A cross-section of the spinal cord shows gray matter (containing cell bodies and interneurons) and white matter (containing axons).

Summary

The vertebrate central nervous system contains the brain and the spinal cord, which are covered and protected by three meninges. The brain contains structurally and functionally defined regions. In mammals, these include the cortex (which can be broken down into four primary functional lobes: frontal, temporal, occipital, and parietal), basal ganglia, thalamus, hypothalamus, limbic system, cerebellum, and brainstem—although structures in some of these designations overlap. While functions may be primarily localized to one structure in the brain, most complex functions, like language and sleep, involve neurons in multiple brain regions. The spinal cord is the information superhighway that connects the brain with the rest of the body through its connections with peripheral nerves. It transmits sensory and motor input and also controls motor reflexes.

Exercises

1. The _____ lobe contains the visual cortex.
    1. frontal
    2. parietal
    3. temporal
    4. occipital
2. The _____ connects the two cerebral hemispheres.
    1. limbic system
    2. corpus callosum

3. cerebellum
4. pituitary

3. Neurons in the _____ control motor reflexes.
    1. thalmus
    2. spinal cord
    3. parietal lobe
    4. hippocampus
4. What methods can be used to determine the function of a particular brain region?
5. What are the main functions of the spinal cord?

Answers

1. D
2. B
3. B
4. To determine the function of a specific brain area, scientists can look at patients who have damage in that brain area and see what symptoms they exhibit. Researchers can disable the brain structure temporarily using transcranial magnetic stimulation. They can disable or remove the area in an animal model. fMRI can be used to correlate specific functions with increased blood flow to brain regions.
5. The spinal cord transmits sensory information from the body to the brain and motor commands from the brain to the body through its connections with peripheral nerves. It also controls motor reflexes.

Glossary

amygdala
structure within the limbic system that processes fear

arachnoid mater
spiderweb-like middle layer of the meninges that cover the central nervous system

basal ganglia
interconnected collections of cells in the brain that are involved in movement and motivation; also known as basal nuclei

brainstem
portion of the brain that connects with the spinal cord; controls basic nervous system functions like breathing, heart rate, and swallowing

cerebellum
brain structure involved in posture, motor coordination, and learning new motor actions

cerebral cortex
outermost sheet of brain tissue; involved in many higher-order functions

cerebrospinal fluid (CSF)
clear liquid that surrounds the brain and spinal cord and fills the ventricles and central canal; acts as a shock absorber and circulates material throughout the brain and spinal cord.

choroid plexus
spongy tissue within ventricles that produces cerebrospinal fluid

cingulate gyrus
helps regulate emotions and pain; thought to directly drive the body's conscious response to unpleasant experiences

corpus callosum
thick fiber bundle that connects the cerebral hemispheres

dura mater
tough outermost layer that covers the central nervous system
depolarization of a postsynaptic membrane caused by neurotransmitter molecules released from a presynaptic cell

frontal lobe
part of the cerebral cortex that contains the motor cortex and areas involved in planning, attention, and language

gyrus
(plural: gyri) ridged protrusions in the cortex

hippocampus
brain structure in the temporal lobe involved in processing memories

hypothalamus
brain structure that controls hormone release and body homeostasis

limbic system
connected brain areas that process emotion and motivation

meninge
membrane that covers and protects the central nervous system

occipital lobe
part of the cerebral cortex that contains visual cortex and processes visual stimuli

parietal lobe
part of the cerebral cortex involved in processing touch and the sense of the body in space

pia mater
thin membrane layer directly covering the brain and spinal cord

proprioception
sense about how parts of the body are oriented in space

somatosensation
sense of touch

sulcus
(plural: sulci) indents or “valleys” in the cortex

temporal lobe
part of the cerebral cortex that processes auditory input; parts of the temporal lobe are involved in speech, memory, and emotion processing

thalamus
brain area that relays sensory information to the cortex

16.4 The Peripheral Nervous System

Learning Objectives

By the end of this section, you will be able to:

- Describe the organization and functions of the sympathetic and parasympathetic nervous systems
- Describe the organization and function of the sensory-somatic 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 can be broken down into the **autonomic nervous system**, which controls bodily functions without conscious control, and the **sensory-somatic nervous system**, which transmits sensory information from the skin, muscles, and sensory organs to the CNS and sends motor commands from the CNS to the muscles.

Autonomic Nervous System

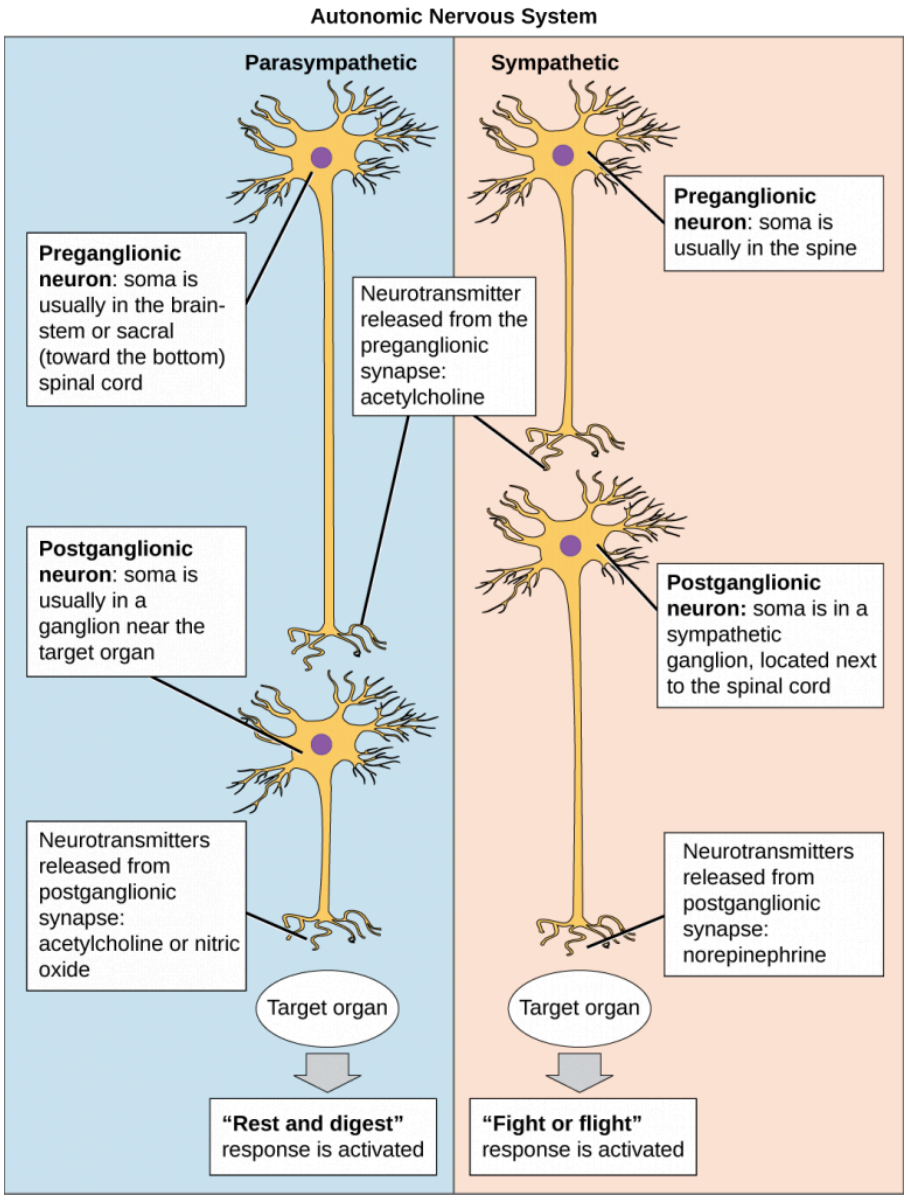


Figure 16.26. In the autonomic nervous system, a preganglionic neuron of the CNS synapses with a postganglionic neuron of the PNS. The postganglionic neuron, in turn, acts on a target organ. Autonomic responses are mediated by the sympathetic and the parasympathetic systems, which are antagonistic to one another. The sympathetic system activates the "fight or flight" response, while the parasympathetic system activates the "rest and digest" response.

Which of the following statements is false?

1. The parasympathetic pathway is responsible for resting the body, while the sympathetic pathway is responsible for preparing for an emergency.
2. Most preganglionic neurons in the sympathetic pathway originate in the spinal cord.
3. Slowing of the heartbeat is a parasympathetic response.
4. Parasympathetic neurons are responsible for releasing norepinephrine on the target organ, while sympathetic neurons are responsible for releasing acetylcholine.

The autonomic nervous system serves as the relay between the CNS and the internal organs. It controls the lungs, the heart, smooth muscle, and exocrine and endocrine glands. The autonomic nervous system controls these organs largely without conscious control; it can continuously monitor the conditions of these different systems and implement changes as needed. Signaling to the target tissue usually involves two synapses: a preganglionic neuron (originating in the CNS) synapses to a neuron in a ganglion that, in turn, synapses on the target organ, as illustrated in Figure 16.26. There are two divisions of the autonomic nervous system that often have opposing effects: the sympathetic nervous system and the parasympathetic nervous system.

Sympathetic Nervous System

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

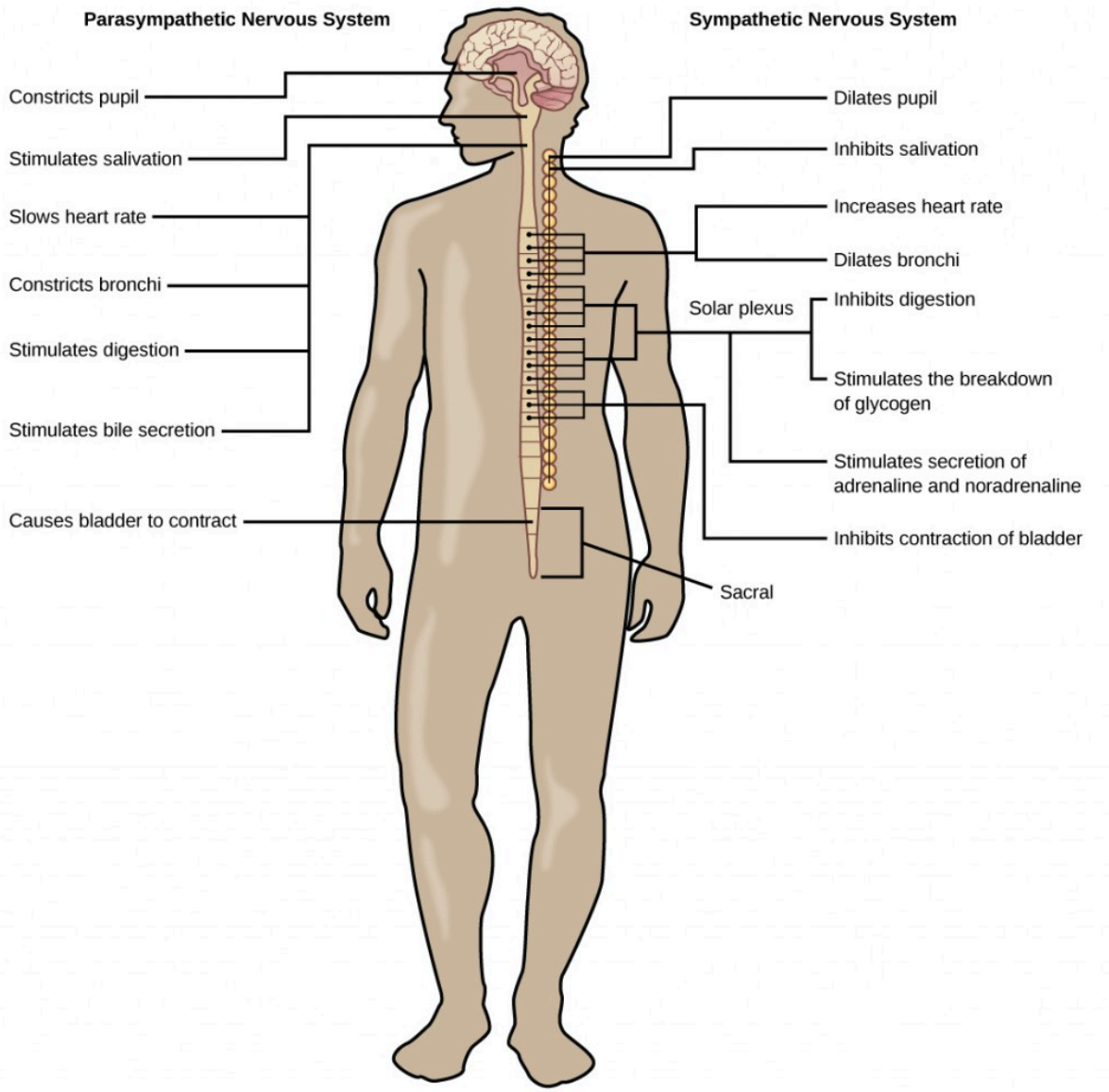


Figure 16.27. The sympathetic and parasympathetic nervous systems often have opposing effects on target organs.

Most preganglionic neurons in the sympathetic nervous system originate in the spinal cord, as illustrated in Figure 16.27. 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 nervous system are quite pervasive. This is both because one preganglionic neuron synapses on multiple postganglionic neurons, amplifying the effect of the original synapse, and because the adrenal gland also releases norepinephrine (and the closely related hormone epinephrine) into the 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, and

scientists have found evidence that it may also increase LTP—allowing the animal to remember the dangerous situation and avoid it in the future.

Parasympathetic Nervous System

While the sympathetic nervous system is activated in stressful situations, the **parasympathetic nervous system** allows an animal to "rest and digest." One way to remember this is to think that during a restful situation like a picnic, the parasympathetic nervous system is in control ("picnic" and "parasympathetic" both start with "p"). Parasympathetic preganglionic neurons have cell bodies located in the brainstem and in the sacral (toward the bottom) spinal cord, as shown in Figure 16.27. The axons of the preganglionic neurons release acetylcholine on the postganglionic neurons, which are generally located very near the target organs. Most postganglionic neurons release acetylcholine onto target organs, although some release nitric oxide.

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

Sensory-Somatic Nervous System

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

Humans have 12 **cranial nerves**, nerves that emerge from or enter the skull (cranium), as opposed to the spinal nerves, which emerge from the vertebral column. Each cranial nerve is accorded a name, which are detailed in Figure 16.28. 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).

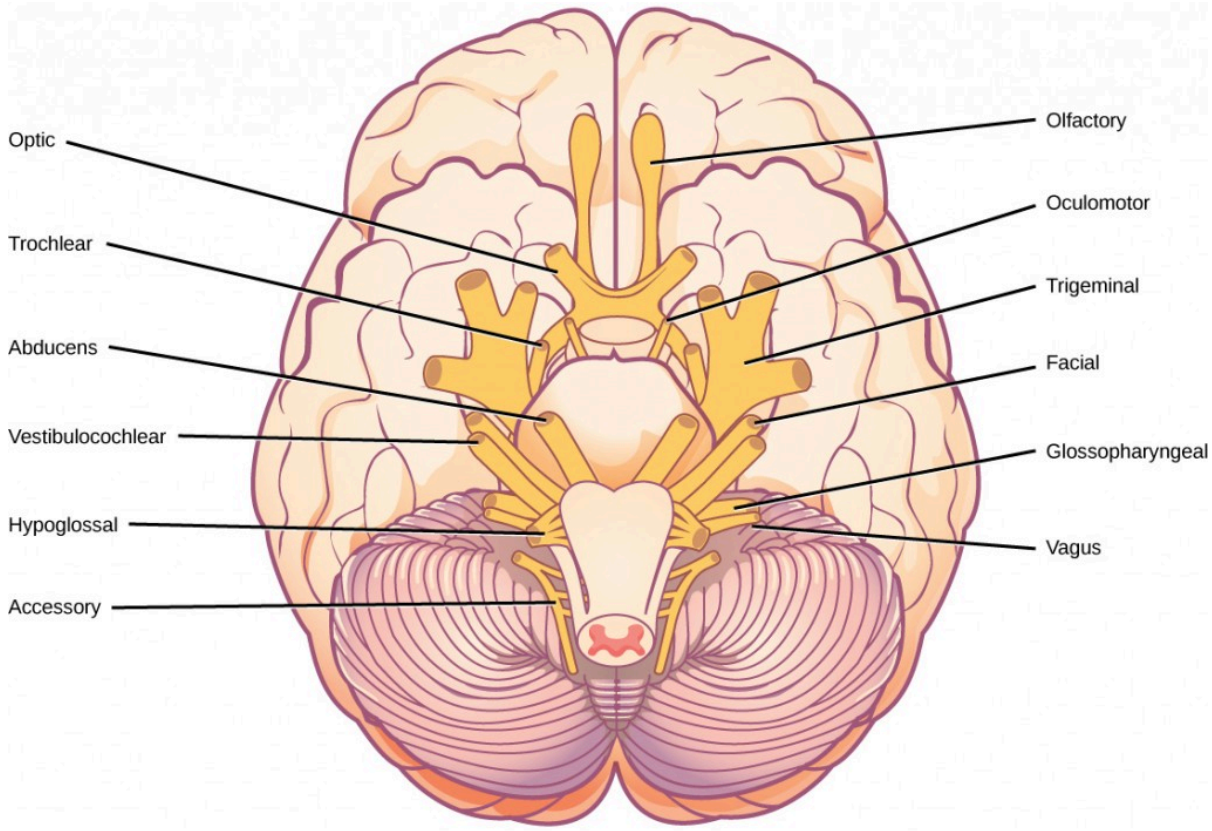


Figure 16.28. The human brain contains 12 cranial nerves that receive sensory input and control motor output for the head and neck.

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 16.29. 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.

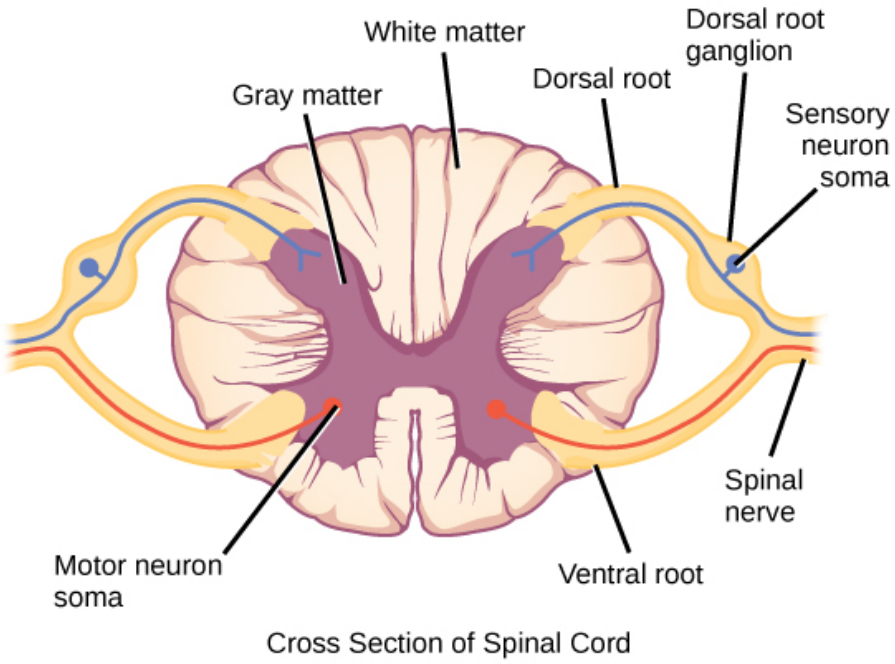


Figure 16.29. Spinal nerves contain both sensory and motor axons. The somas of sensory neurons are located in dorsal root ganglia. The somas of motor neurons are found in the ventral portion of the gray matter of the spinal cord.

Summary

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

Exercises

1. Which of the following statements is false?
    1. The parasympathetic pathway is responsible for relaxing the body, while the sympathetic pathway is responsible for preparing for an emergency.
    2. Most preganglionic neurons in the sympathetic pathway originate in the spinal cord.
    3. Slowing of the heartbeat is a parasympathetic response.
    4. Parasympathetic neurons are responsible for releasing norepinephrine on the target organ, while sympathetic neurons are responsible for releasing acetylcholine.
2. Activation of the sympathetic nervous system causes:

1. increased blood flow into the skin
2. a decreased heart rate
3. an increased heart rate
4. increased digestion

3. Where are parasympathetic preganglionic cell bodies located?

    1. cerebellum
    2. brainstem
    3. dorsal root ganglia
    4. skin

4. _____ is released by motor nerve endings onto muscle.

    1. Acetylcholine
    2. Norepinephrine
    3. Dopamine
    4. Serotonin

5. What are the main differences between the sympathetic and parasympathetic branches of the autonomic nervous system?
6. What are the main functions of the sensory-somatic nervous system?

Answers

1. D
2. C
3. B
4. A
5. The sympathetic nervous system prepares the body for “fight or flight,” whereas the parasympathetic nervous system allows the body to “rest and digest.” Sympathetic neurons release norepinephrine onto target organs; parasympathetic neurons release acetylcholine. Sympathetic neuron cell bodies are located in sympathetic ganglia. Parasympathetic neuron cell bodies are located in the brainstem and sacral spinal cord. Activation of the sympathetic nervous system increases heart rate and blood pressure and decreases digestion and blood flow to the skin. Activation of the parasympathetic nervous system decreases heart rate and blood pressure and increases digestion and blood flow to the skin.
6. The sensory-somatic nervous system transmits sensory information from the skin, muscles, and sensory organs to the CNS. It also sends motor commands from the CNS to the muscles, causing them to contract.

Glossary

acetylcholine
neurotransmitter released by neurons in the central nervous system and peripheral nervous system

autonomic nervous system
part of the peripheral nervous system that controls bodily functions

cranial nerve
sensory and/or motor nerve that emanates from the brain

norepinephrine
neurotransmitter and hormone released by activation of the sympathetic nervous system

parasympathetic nervous system
division of autonomic nervous system that regulates visceral functions during rest and digestion

sensory-somatic nervous system
system of sensory and motor nerves

spinal nerve
nerve projecting between skin or muscle and spinal cord

sympathetic nervous system
division of autonomic nervous system activated during stressful “fight or flight” situations

16.5 Nervous System Disorders

Learning Objectives

By the end of this section, you will be able to:

- Describe the symptoms, potential causes, and treatment of several examples of nervous system disorders

A nervous system that functions correctly is a fantastically complex, well-oiled machine—synapses fire appropriately, muscles move when needed, memories are formed and stored, and emotions are well regulated. Unfortunately, each year millions of people in the United States deal with some sort of nervous system disorder. While scientists have discovered potential causes of many of these diseases, and viable treatments for some, ongoing research seeks to find ways to better prevent and treat all of these disorders.

Neurodegenerative Disorders

Neurodegenerative disorders are illnesses characterized by a loss of nervous system functioning that are usually caused by neuronal death. These diseases generally worsen over time as more and more neurons die. The symptoms of a particular neurodegenerative disease are related to where in the nervous system the death of neurons occurs. Spinocerebellar ataxia, for example, leads to neuronal death in the cerebellum. The death of these neurons causes problems in balance and walking. Neurodegenerative disorders include Huntington's disease, amyotrophic lateral sclerosis, Alzheimer's disease and other types of dementia disorders, and Parkinson's disease. Here, Alzheimer's and Parkinson's disease will be discussed in more depth.

Alzheimer's Disease

Alzheimer's disease is the most common cause of dementia in the elderly. In 2012, an estimated 5.4 million Americans suffered from Alzheimer's disease, and payments for their care are estimated at $200 billion. Roughly one in every eight people age 65 or older has the disease. Due to the aging of the baby-boomer generation, there are projected to be as many as 13 million Alzheimer's patients in the United States in the year 2050.

Symptoms of Alzheimer's disease include disruptive memory loss, confusion about time or place, difficulty planning or executing tasks, poor judgment, and personality changes. Problems smelling certain scents can also be indicative of Alzheimer's disease and may serve as an early warning sign. Many of these symptoms are also common in people who are aging normally, so it is the severity and longevity of the symptoms that determine whether a person is suffering from Alzheimer's.

Alzheimer's disease was named for Alois Alzheimer, a German psychiatrist who published a report in 1911 about a woman who showed severe dementia symptoms. Along with his colleagues, he examined the woman's brain following her death and reported the presence of abnormal clumps, which are now called amyloid plaques, along with tangled brain fibers called neurofibrillary tangles. Amyloid plaques, neurofibrillary tangles, and an overall shrinking of brain volume are commonly seen in the brains of Alzheimer's patients. Loss of neurons in the hippocampus is especially severe in advanced Alzheimer's patients. Figure 16.30 compares a normal brain to the brain of an Alzheimer's patient. Many research groups are examining the causes of these hallmarks of the disease.

One form of the disease is usually caused by mutations in one of three known genes. This rare form of early onset Alzheimer's disease affects fewer than five percent of patients with the disease and causes dementia beginning between the ages of 30 and 60. The more prevalent, late-onset form of the disease likely also has a genetic component. One particular gene, apolipoprotein E (APOE) has a variant (E4) that increases a carrier's likelihood of getting the disease. Many other genes have been identified that might be involved in the pathology.

Concept in Action

Visit this website for video links discussing genetics and Alzheimer's disease.

Unfortunately, there is no cure for Alzheimer's disease. Current treatments focus on managing the symptoms of the disease. Because decrease in the activity of cholinergic neurons (neurons that use the neurotransmitter acetylcholine) is common in Alzheimer's disease, several drugs used to treat the disease work by increasing acetylcholine neurotransmission, often by inhibiting the enzyme that breaks down acetylcholine in the synaptic cleft. Other clinical interventions focus on behavioral therapies like psychotherapy, sensory therapy, and cognitive exercises. Since Alzheimer's disease appears to hijack the normal aging process, research into prevention is prevalent. Smoking, obesity, and cardiovascular problems may be risk factors for the disease, so treatments for those may also help to prevent Alzheimer's disease. Some studies have shown that people who remain intellectually active by playing games, reading, playing musical instruments, and being socially active in later life have a reduced risk of developing the disease.

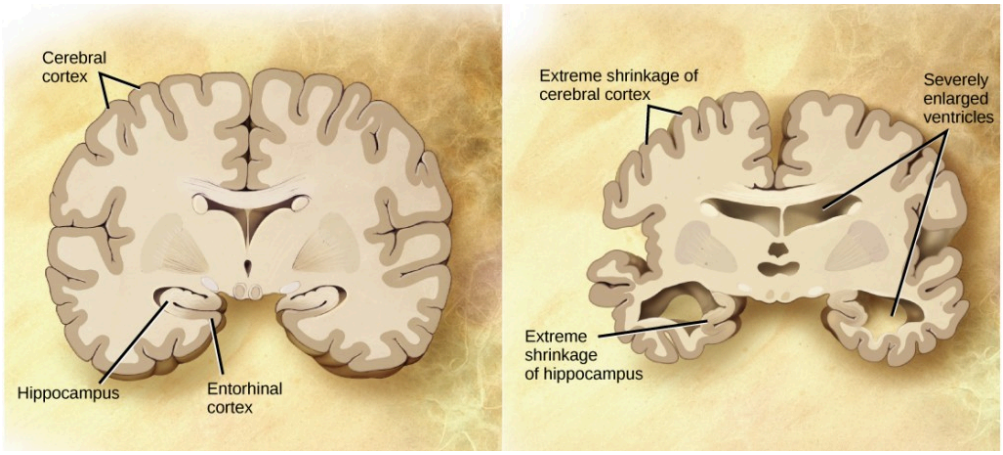


Figure 16.30. Compared to a normal brain (left), the brain from a patient with Alzheimer's disease (right) shows a dramatic neurodegeneration, particularly within the ventricles and hippocampus. (credit: modification of work by "Garrando"/Wikimedia Commons based on original images by ADEAR: "Alzheimer's Disease Education and Referral Center, a service of the National Institute on Aging")

Parkinson's Disease

Like Alzheimer's disease, **Parkinson's disease** is a neurodegenerative disease. It was first characterized by James Parkinson in 1817. Each year, 50,000-60,000 people in the United States are diagnosed with the disease. Parkinson's disease causes the loss of dopamine neurons in the substantia nigra, a midbrain structure that regulates movement. Loss of these neurons causes many symptoms including tremor (shaking of fingers or a limb), slowed movement, speech changes, balance and posture problems, and rigid muscles. The combination of these symptoms often causes a characteristic slow hunched shuffling walk, illustrated in Figure 16.31. Patients with Parkinson's disease can also exhibit psychological symptoms, such as dementia or emotional problems.

Although some patients have a form of the disease known to be caused by a single mutation, for most patients the exact causes of Parkinson's disease remain unknown: the disease likely results from a combination of genetic and environmental factors (similar to Alzheimer's disease). Post-mortem analysis of brains from Parkinson's patients shows the presence of Lewy bodies—abnormal protein clumps—in dopaminergic neurons. The prevalence of these Lewy bodies often correlates with the severity of the disease.

There is no cure for Parkinson's disease, and treatment is focused on easing symptoms. One of the most commonly prescribed drugs for Parkinson's is L-DOPA, which is a chemical that is converted into dopamine by neurons in the brain. This conversion increases the overall level of dopamine neurotransmission and can help compensate for the loss of dopaminergic neurons in the substantia nigra. Other drugs work by inhibiting the enzyme that breaks down dopamine.

Figure 16.31. Parkinson's patients often have a characteristic hunched walk.

Neurodevelopmental Disorders

Neurodevelopmental disorders occur when the development of the nervous system is disturbed. There are several different classes of neurodevelopmental disorders. Some, like Down Syndrome, cause intellectual deficits. Others specifically affect communication, learning, or the motor system. Some disorders like autism spectrum disorder and attention deficit/hyperactivity disorder have complex symptoms.

Autism

Autism spectrum disorder (ASD) is a neurodevelopmental disorder. Its severity differs from person to person. Estimates for the prevalence of the disorder have changed rapidly in the past few decades. Current estimates suggest that one in 88 children will develop the disorder. ASD is four times more prevalent in males than females.

Concept in Action

This video discusses possible reasons why there has been a recent increase in the number of people diagnosed with autism.

A characteristic symptom of ASD is impaired social skills. Children with autism may have difficulty making and maintaining eye contact and reading social cues. They also may have problems feeling empathy for others. Other symptoms of ASD include repetitive motor behaviors (such as rocking back and forth), preoccupation with specific subjects, strict adherence to certain rituals, and unusual language use. Up to 30 percent of patients with ASD develop epilepsy, and patients with some forms of the disorder (like Fragile X) also have intellectual disability. Because it is a spectrum disorder, other ASD patients are very functional and have good-to-excellent language skills. Many of these patients do not feel that they suffer from a disorder and instead think that their brains just process information differently.

Except for some well-characterized, clearly genetic forms of autism (like Fragile X and Rett's Syndrome), the causes of ASD are largely unknown. Variants of several genes correlate with the presence of ASD, but for any given patient, many different mutations in different genes may be required for the disease to develop. At a general level, ASD is thought to be a disease of "incorrect" wiring. Accordingly, brains of some ASD patients lack the same level of synaptic pruning that occurs in non-affected people. In the 1990s, a research paper linked autism to a common vaccine given to children. This paper was retracted when it was discovered that the author falsified data, and follow-up studies showed no connection between vaccines and autism.

Treatment for autism usually combines behavioral therapies and interventions, along with medications to treat other disorders common to people with autism (depression, anxiety, obsessive compulsive disorder). Although early interventions can help mitigate the effects of the disease, there is currently no cure for ASD.

Attention Deficit Hyperactivity Disorder (ADHD)

Approximately three to five percent of children and adults are affected by **attention deficit/ hyperactivity disorder (ADHD)**. Like ASD, ADHD is more prevalent in males than females. Symptoms of the disorder include inattention (lack of focus), executive functioning difficulties, impulsivity, and hyperactivity beyond what is characteristic of the normal developmental stage. Some patients do not have the hyperactive component of symptoms and are diagnosed with a subtype of ADHD: attention deficit disorder (ADD). Many people with ADHD also show comorbitity, in that they develop secondary disorders in addition to ADHD. Examples include depression or obsessive

compulsive disorder (OCD). Figure 16.32 provides some statistics concerning comorbidity with ADHD.

The cause of ADHD is unknown, although research points to a delay and dysfunction in the development of the prefrontal cortex and disturbances in neurotransmission. According to studies of twins, the disorder has a strong genetic component. There are several candidate genes that may contribute to the disorder, but no definitive links have been discovered. Environmental factors, including exposure to certain pesticides, may also contribute to the development of ADHD in some patients. Treatment for ADHD often involves behavioral therapies and the prescription of stimulant medications, which paradoxically cause a calming effect in these patients.

Figure 16.32. Many people with ADHD have one or more other neurological disorders. (credit "chart design and illustration": modification of work by Leigh Coriale; credit "data": Drs. Biederman and Faraone, Massachusetts General Hospital).

Neurologist

Neurologists are physicians who specialize in disorders of the nervous system. They diagnose and treat disorders such as epilepsy, stroke, dementia, nervous system injuries, Parkinson's disease, sleep disorders, and multiple sclerosis. Neurologists are medical doctors who have attended college, medical school, and completed three to four years of neurology residency.

When examining a new patient, a neurologist takes a full medical history and performs a complete physical exam. The physical exam contains specific tasks that are used to determine what areas of the brain, spinal cord, or peripheral nervous system may be damaged. For example, to check whether the hypoglossal nerve is functioning correctly, the neurologist will ask the patient to move his or her tongue in different ways. If the patient does not have full control over tongue movements, then the hypoglossal nerve may be damaged or there may be a lesion in the brainstem where the cell bodies of these neurons reside (or there could be damage to the tongue muscle itself).

Neurologists have other tools besides a physical exam they can use to diagnose particular problems in the nervous system. If the patient has had a seizure, for example, the neurologist can use electroencephalography (EEG), which involves taping electrodes to the scalp to record brain activity, to try to determine which brain regions are involved in the seizure. In suspected stroke patients, a neurologist can use a computerized tomography (CT) scan, which is a type of X-ray, to look for

bleeding in the brain or a possible brain tumor. To treat patients with neurological problems, neurologists can prescribe medications or refer the patient to a neurosurgeon for surgery.

Concept in Action

This website allows you to see the different tests a neurologist might use to see what regions of the nervous system may be damaged in a patient.

Mental Illnesses

Mental illnesses are nervous system disorders that result in problems with thinking, mood, or relating with other people. These disorders are severe enough to affect a person's quality of life and often make it difficult for people to perform the routine tasks of daily living. Debilitating mental disorders plague approximately 12.5 million Americans (about 1 in 17 people) at an annual cost of more than $300 billion. There are several types of mental disorders including schizophrenia, major depression, bipolar disorder, anxiety disorders and phobias, post-traumatic stress disorders, and obsessive-compulsive disorder (OCD), among others. The American Psychiatric Association publishes the Diagnostic and Statistical Manual of Mental Disorders (or DSM), which describes the symptoms required for a patient to be diagnosed with a particular mental disorder. Each newly released version of the DSM contains different symptoms and classifications as scientists learn more about these disorders, their causes, and how they relate to each other. A more detailed discussion of two mental illnesses—schizophrenia and major depression—is given below.

Schizophrenia

Schizophrenia is a serious and often debilitating mental illness affecting one percent of people in the United States. Symptoms of the disease include the inability to differentiate between reality and imagination, inappropriate and unregulated emotional responses, difficulty thinking, and problems with social situations. People with schizophrenia can suffer from hallucinations and hear voices; they may also suffer from delusions. Patients also have so-called "negative" symptoms like a flattened emotional state, loss of pleasure, and loss of basic drives. Many schizophrenic patients are diagnosed in their late adolescence or early 20s. The development of schizophrenia is thought to involve malfunctioning dopaminergic neurons and may also involve problems with glutamate signaling. Treatment for the disease usually requires antipsychotic medications that work by blocking dopamine receptors and decreasing dopamine neurotransmission in the brain. This decrease in dopamine can cause Parkinson's disease-like symptoms in some patients. While some classes of antipsychotics can be quite effective at treating the disease, they are not a cure, and most patients must remain medicated for the rest of their lives.

Depression

Major depression affects approximately 6.7 percent of the adults in the United States each year and is one of the most common mental disorders. To be diagnosed with major depressive disorder, a person must have experienced a severely depressed mood lasting longer than two weeks along with other symptoms including a loss of enjoyment in activities that were previously enjoyed, changes in appetite and sleep schedules, difficulty concentrating, feelings of worthlessness, and suicidal thoughts. The exact causes of major depression are unknown and likely include both genetic and environmental risk factors. Some research supports the "classic monoamine hypothesis," which suggests that depression is caused by a decrease in norepinephrine and serotonin neurotransmission. One argument against this hypothesis is the fact that some antidepressant medications cause an increase in norepinephrine and serotonin release within a few hours of beginning treatment—but clinical results of these medications are not seen until weeks later. This has led to alternative hypotheses: for example, dopamine may also be decreased in depressed patients, or it may actually be an increase in norepinephrine and serotonin that causes the disease, and antidepressants force a feedback loop that decreases this release. Treatments for depression include psychotherapy, electroconvulsive therapy, deep-brain stimulation, and prescription medications. There are several classes of antidepressant medications that work through different mechanisms. For example, monoamine oxidase inhibitors (MAO inhibitors) block the enzyme that degrades many neurotransmitters (including dopamine, serotonin, norepinephrine), resulting in increased neurotransmitter in the synaptic cleft. Selective serotonin reuptake inhibitors (SSRIs) block the reuptake of serotonin into the presynaptic neuron. This blockage results in an increase in serotonin in the synaptic cleft. Other types of drugs such as norepinephrine-dopamine reuptake inhibitors and norepinephrine-serotonin reuptake inhibitors are also used to treat depression.

Other Neurological Disorders

There are several other neurological disorders that cannot be easily placed in the above categories. These include chronic pain conditions, cancers of the nervous system, epilepsy disorders, and stroke. Epilepsy and stroke are discussed below.

Epilepsy

Estimates suggest that up to three percent of people in the United States will be diagnosed with **epilepsy** in their lifetime. While there are several different types of epilepsy, all are characterized by recurrent seizures. Epilepsy itself can be a symptom of a brain injury, disease, or other illness. For example, people who have intellectual disability or ASD can experience seizures, presumably because the developmental wiring malfunctions that caused their disorders also put them at risk for epilepsy. For many patients, however, the cause of their epilepsy is never identified and is likely to be a combination of genetic and environmental factors. Often, seizures can be controlled with anticonvulsant medications. However, for very severe cases, patients may undergo brain surgery to remove the brain area where seizures originate.

Stroke

A stroke results when blood fails to reach a portion of the brain for a long enough time to cause

damage. Without the oxygen supplied by blood flow, neurons in this brain region die. This neuronal death can cause many different symptoms—depending on the brain area affected— including headache, muscle weakness or paralysis, speech disturbances, sensory problems, memory loss, and confusion. Stroke is often caused by blood clots and can also be caused by the bursting of a weak blood vessel. Strokes are extremely common and are the third most common cause of death in the United States. On average one person experiences a stroke every 40 seconds in the United States. Approximately 75 percent of strokes occur in people older than 65. Risk factors for stroke include high blood pressure, diabetes, high cholesterol, and a family history of stroke. Smoking doubles the risk of stroke. Because a stroke is a medical emergency, patients with symptoms of a stroke should immediately go to the emergency room, where they can receive drugs that will dissolve any clot that may have formed. These drugs will not work if the stroke was caused by a burst blood vessel or if the stroke occurred more than three hours before arriving at the hospital. Treatment following a stroke can include blood pressure medication (to prevent future strokes) and (sometimes intense) physical therapy.

Summary

Some general themes emerge from the sampling of nervous system disorders presented above. The causes for most disorders are not fully understood—at least not for all patients—and likely involve a combination of nature (genetic mutations that become risk factors) and nurture (emotional trauma, stress, hazardous chemical exposure). Because the causes have yet to be fully determined, treatment options are often lacking and only address symptoms.

Exercises

1. Parkinson’s disease is a caused by the degeneration of neurons that release _____.
    1. serotonin
    2. dopamine
    3. glutamate
    4. norepinephrine
2. _____ medications are often used to treat patients with ADHD.
    1. Tranquilizer
    2. Antibiotic
    3. Stimulant
    4. Anti-seizure
3. Strokes are often caused by _____.
    1. neurodegeneration
    2. blood clots or burst blood vessels
    3. seizures
    4. viruses

4. What are the main symptoms of Alzheimer's disease?
5. What are possible treatments for patients with major depression?

Answers

1. B
2. C
3. B
4. Symptoms of Alzheimer's disease include disruptive memory loss, confusion about time or place, difficulties planning or executing tasks, poor judgment, and personality changes.
5. Possible treatments for patients with major depression include psychotherapy and prescription medications. MAO inhibitor drugs inhibit the breakdown of certain neurotransmitters (including dopamine, serotonin, norepinephrine) in the synaptic cleft. SSRI medications inhibit the reuptake of serotonin into the presynaptic neuron.

Glossary

Alzheimer's disease
neurodegenerative disorder characterized by problems with memory and thinking

attention deficit hyperactivity disorder (ADHD)
neurodevelopmental disorder characterized by difficulty maintaining attention and controlling impulses

autism spectrum disorder (ASD)
neurodevelopmental disorder characterized by impaired social interaction and communication abilities

epilepsy
neurological disorder characterized by recurrent seizures

major depression
mental illness characterized by prolonged periods of sadness

neurodegenerative disorder
nervous system disorder characterized by the progressive loss of neurological functioning, usually caused by neuron death

Parkinson's disease
neurodegenerative disorder that affects the control of movement

schizophrenia
mental disorder characterized by the inability to accurately perceive reality; patients often have difficulty thinking clearly and can suffer from delusions

Chapter 17. Sensory Systems

Figure 17.1. This shark uses its senses of sight, vibration (lateral-line system), and smell to hunt, but it also relies on its ability to sense the electric fields of prey, a sense not present in most land animals. (credit: modification of work by Hermanus Backpackers Hostel, South Africa)

Introduction

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

17.1 Sensory Processes

Learning Objectives

By the end of this section, you will be able to:

- Identify the general and special senses in humans
- Describe three important steps in sensory perception
- Explain the concept of just-noticeable difference in sensory perception

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

There are two broad types of cellular systems that perform sensory transduction. In one, a neuron works with a **sensory receptor**, a cell, or cell process that is specialized to engage with and detect a specific stimulus. Stimulation of the sensory receptor activates the associated afferent neuron, which carries information about the stimulus to the central nervous system. In the second type of sensory transduction, a sensory nerve ending responds to a stimulus in the internal or external environment: this neuron constitutes the sensory receptor. Free nerve endings can be stimulated by several different stimuli, thus showing little receptor specificity. For example, pain receptors in your gums and teeth may be stimulated by temperature changes, chemical stimulation, or pressure.

Reception

The first step in sensation is **reception**

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

Transduction

The most fundamental function of a sensory system is the translation of a sensory signal to an electrical signal in the nervous system. This takes place at the sensory receptor, and the change in electrical potential that is produced is called the **receptor potential**. How is sensory input, such as pressure on the skin, changed to a receptor potential? In this example, a type of receptor called a **mechanoreceptor** (as shown in

Figure 17.2) possesses specialized membranes that respond to pressure. Disturbance of these dendrites by compressing them or bending them opens gated ion channels in the plasma membrane of the sensory neuron, changing its electrical potential. Recall that in the nervous system, a positive change of a neuron's electrical potential (also called the membrane potential), depolarizes the neuron. Receptor potentials are graded potentials: the magnitude of these graded (receptor) potentials varies with the strength of the stimulus. If the magnitude of depolarization is sufficient (that is, if membrane potential reaches a threshold), the neuron will fire an action potential. In most cases, the correct stimulus impinging on a sensory receptor will drive membrane potential in a positive direction, although for some receptors, such as those in the visual system, this is not always the case.

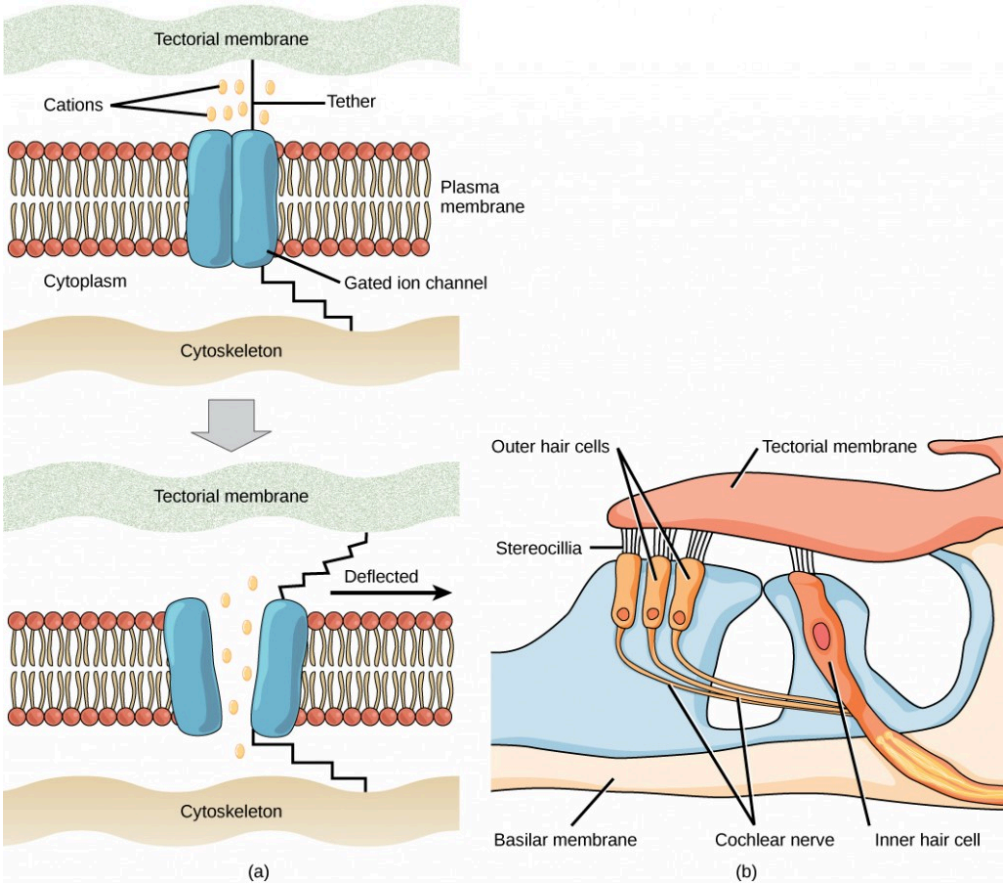


Figure 17.2. (a) Mechanosensitive ion channels are gated ion channels that respond to mechanical deformation of the plasma membrane. A mechanosensitive channel is connected to the plasma membrane and the cytoskeleton by hair-like tethers. When pressure causes the extracellular matrix to move, the channel opens, allowing ions to enter or exit the cell. (b) Stereocilia in the human ear are connected to mechanosensitive ion channels. When a sound causes the stereocilia to move, mechanosensitive ion channels transduce the signal to the cochlear nerve.

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

Encoding and Transmission of Sensory Information

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

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

Perception

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

All sensory signals, except those from the olfactory system, are transmitted though the central nervous system and are routed to the thalamus and to the appropriate region of the cortex. Recall that the thalamus is a structure in the forebrain that serves as a clearinghouse and relay station for sensory (as well as motor) signals. When the sensory signal exits the thalamus, it is conducted to the specific area of the cortex (Figure 17.3) dedicated to processing that particular sense.

How are neural signals interpreted? Interpretation of sensory signals between individuals of the same species is largely similar, owing to the inherited similarity of their nervous systems; however, there are some individual differences. A good example of this is individual tolerances to a painful stimulus, such as dental pain, which certainly differ.

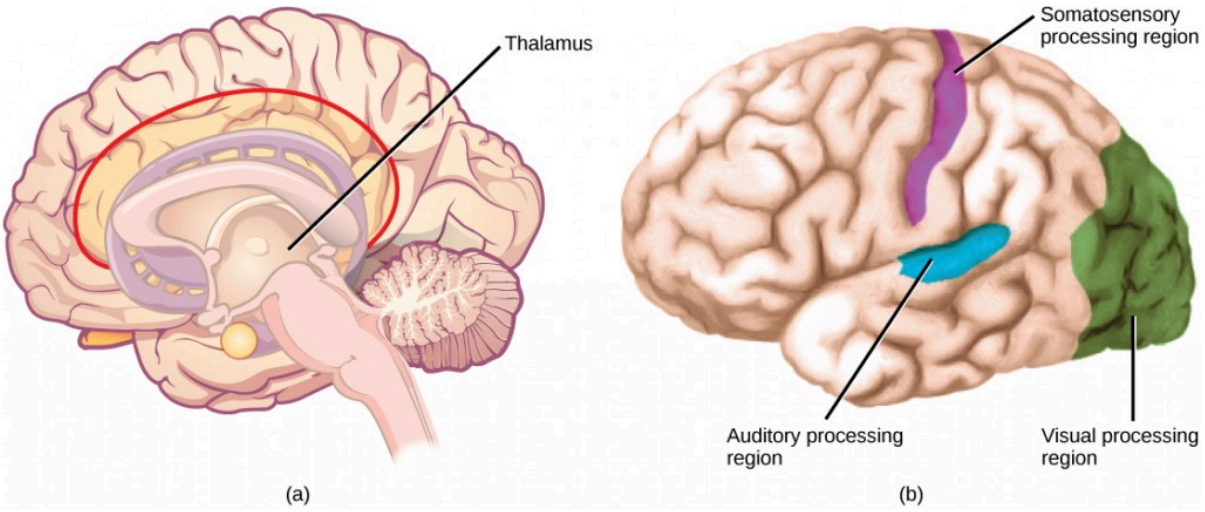


Figure 17.3. In humans, with the exception of olfaction, all sensory signals are routed from the (a) thalamus to (b) final processing regions in the cortex of the brain. (credit b: modification of work by Polina Tishina) Scientific Method Connection

Just-Noticeable Difference

It is easy to differentiate between a one-pound bag of rice and a two-pound bag of rice. There is a one-pound difference, and one bag is twice as heavy as the other. However, would it be as easy to differentiate between a 20- and a 21-pound bag?

Question: What is the smallest detectible weight difference between a one-pound bag of rice and a larger bag? What is the smallest detectible difference between a 20-pound bag and a larger bag? In both cases, at what weights are the differences detected? This smallest detectible difference in stimuli is known as the just-noticeable difference (JND).

Background: Research background literature on JND and on Weber's Law, a description of a proposed mathematical relationship between the overall magnitude of the stimulus and the JND. You will be testing JND of different weights of rice in bags. Choose a convenient increment that is to be stepped through while testing. For example, you could choose 10 percent increments between one and two pounds (1.1, 1.2, 1.3, 1.4, and so on) or 20 percent increments (1.2, 1.4, 1.6, and 1.8).

Hypothesis: Develop a hypothesis about JND in terms of percentage of the whole weight being tested (such as "the JND between the two small bags and between the two large bags is proportionally the same," or ". . . is not proportionally the same.") So, for the first hypothesis, if the JND between the one-pound bag and a larger bag is 0.2 pounds (that is, 20 percent; 1.0 pound feels the same as 1.1 pounds, but 1.0 pound feels less than 1.2 pounds), then the JND between the 20-pound bag and a larger bag will also be 20 percent. (So, 20 pounds feels the same as 22 pounds or 23 pounds, but 20 pounds feels less than 24 pounds.)

Test the hypothesis: Enlist 24 participants, and split them into two groups of 12. To set up the demonstration, assuming a 10 percent increment was selected, have the first group be the one-pound group. As a counter-balancing measure against a systematic error, however, six of the first group will compare one pound to two pounds, and step down in weight (1.0 to 2.0, 1.0 to 1.9, and so on.), while the other six will step up (1.0 to 1.1, 1.0 to 1.2, and so on). Apply the same principle to the 20-pound

group (20 to 40, 20 to 38, and so on, and 20 to 22, 20 to 24, and so on). Given the large difference between 20 and 40 pounds, you may wish to use 30 pounds as your larger weight. In any case, use two weights that are easily detectable as different.

Record the observations: Record the data in a table similar to the table below. For the one-pound and 20-pound groups (base weights) record a plus sign (+) for each participant that detects a difference between the base weight and the step weight. Record a minus sign (-) for each participant that finds no difference. If one-tenth steps were not used, then replace the steps in the "Step Weight" columns with the step you are using.

Table 17.1. Results of JND Testing (+ = difference; – = no difference)

Step Weight	One pound	20 pounds	Step Weight
1.1			22
1.2			24
1.3			26
1.4			28
1.5			30
1.6			32
1.7			34
1.8			36
1.9			38
2.0			40

Analyze the data/report the results: What step weight did all participants find to be equal with one-pound base weight? What about the 20-pound group?

Draw a conclusion: Did the data support the hypothesis? Are the final weights proportionally the same? If not, why not? Do the findings adhere to Weber's Law? Weber's Law states that the concept that a just-noticeable difference in a stimulus is proportional to the magnitude of the original stimulus.

Summary

A sensory activation occurs when a physical or chemical stimulus is processed into a neural signal (sensory transduction) by a sensory receptor. Perception is an individual interpretation of a sensation and is a brain function. Humans have special senses: olfaction, gustation, equilibrium, and hearing, plus the general senses of somatosensation.

Sensory receptors are either specialized cells associated with sensory neurons or the specialized ends of sensory neurons that are a part of the peripheral nervous system, and they are used to receive

information about the environment (internal or external). Each sensory receptor is modified for the type of stimulus it detects. For example, neither gustatory receptors nor auditory receptors are sensitive to light. Each sensory receptor is responsive to stimuli within a specific region in space, which is known as that receptor's receptive field. The most fundamental function of a sensory system is the translation of a sensory signal to an electrical signal in the nervous system.

All sensory signals, except those from the olfactory system, enter the central nervous system and are routed to the thalamus. When the sensory signal exits the thalamus, it is conducted to the specific area of the cortex dedicated to processing that particular sense.

Exercises

1. Which of the following statements about mechanoreceptors is false?
    1. Pacini corpuscles are found in both glabrous and hairy skin.
    2. Merkel's disks are abundant on the fingertips and lips.
    3. Ruffini endings are encapsulated mechanoreceptors.
    4. Meissner's corpuscles extend into the lower dermis.
2. Where does perception occur?
    1. spinal cord
    2. cerebral cortex
    3. receptors
    4. thalamus
3. If a person's cold receptors no longer convert cold stimuli into sensory signals, that person has a problem with the process of _____.
    1. reception
    2. transmission
    3. perception
    4. transduction
4. After somatosensory transduction, the sensory signal travels through the brain as a(n) _____ signal.
    1. electrical
    2. pressure
    3. optical
    4. thermal

Answers

1. D

2. B
3. D
4. A

Glossary

kinesthesia
sense of body movement

mechanoreceptor
sensory receptor modified to respond to mechanical disturbance such as being bent, touch, pressure, motion, and sound

perception
individual interpretation of a sensation; a brain function

proprioception
sense of limb position; used to track kinesthesia

reception
receipt of a signal (such as light or sound) by sensory receptors

receptive field
region in space in which a stimulus can activate a given sensory receptor

receptor potential
membrane potential in a sensory receptor in response to detection of a stimulus

sensory receptor
specialized neuron or other cells associated with a neuron that is modified to receive specific sensory input

sensory transduction
conversion of a sensory stimulus into electrical energy in the nervous system by a change in the membrane potential

vestibular sense
sense of spatial orientation and balance

17.2 Somatosensation

Learning Objectives

By the end of this section, you will be able to:

- Describe four important mechanoreceptors in human skin
- Describe the topographical distribution of somatosensory receptors between glabrous and hairy skin
- Explain why the perception of pain is subjective

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

Recall that the epidermis is the outermost layer of skin in mammals. It is relatively thin, is composed of keratin-filled cells, and has no blood supply. The epidermis serves as a barrier to water and to invasion by pathogens. Below this, the much thicker dermis contains blood vessels, sweat glands, hair follicles, lymph vessels, and lipid-secreting sebaceous glands (Figure 17.4). Below the epidermis and dermis is the subcutaneous tissue, or hypodermis, the fatty layer that contains blood vessels, connective tissue, and the axons of sensory neurons. The hypodermis, which holds about 50 percent of the body's fat, attaches the dermis to the bone and muscle, and supplies nerves and blood vessels to the dermis.

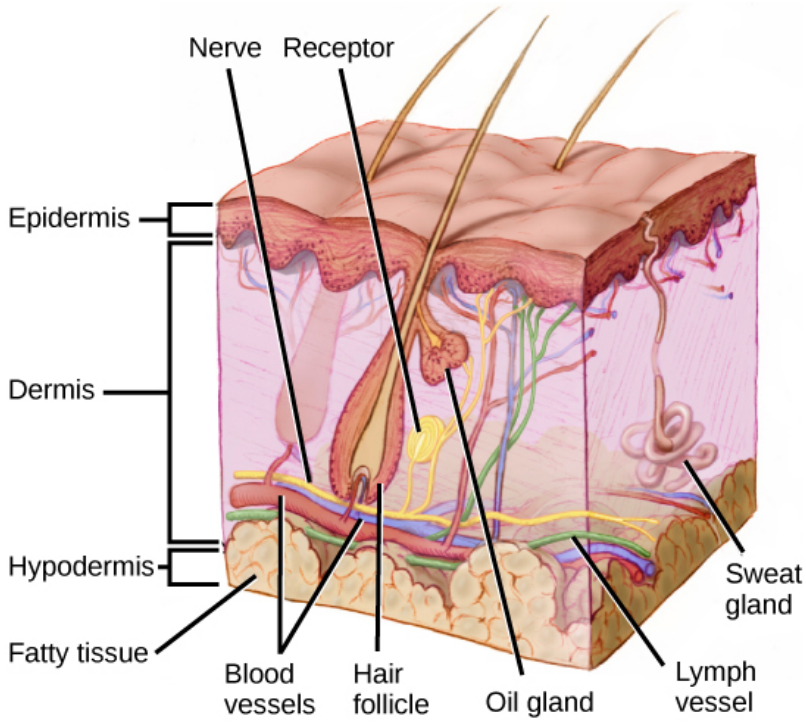


Figure 17.4. Mammalian skin has three layers: an epidermis, a dermis, and a hypodermis. (credit: modification of work by Don Bliss, National Cancer Institute)

Somatosensory Receptors

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

There are three classes of mechanoreceptors: tactile, proprioceptors, and baroreceptors. Mechanoreceptors sense stimuli due to physical deformation of their plasma membranes. They contain mechanically gated ion channels whose gates open or close in response to pressure, touch, stretching, and sound.” There are four primary tactile mechanoreceptors in human skin: Merkel’s disks, Meissner’s corpuscles, Ruffini endings, and Pacinian corpuscle; two are located toward the surface of the skin and two are located deeper. A fifth type of mechanoreceptor, Krause end bulbs, are found only in specialized regions. **Merkel’s disks** (shown in Figure 17.5) are found in the upper layers of skin near the base of the epidermis, both in skin that has hair and on **glabrous** skin, that is, the hairless skin found on the palms and fingers, the soles of the feet, and the lips of humans and other primates. Merkel’s disks are densely distributed in the fingertips and lips. They are slow-adapting, unencapsulated nerve endings, and they respond to light touch. Light touch, also known as discriminative touch, is a light pressure that allows the location of a stimulus to be pinpointed. The

receptive fields of Merkel's disks are small with well-defined borders. That makes them finely sensitive to edges and they come into use in tasks such as typing on a keyboard.

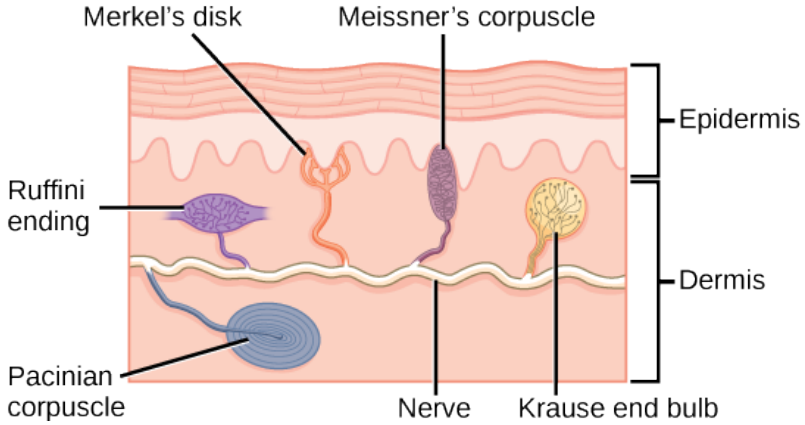


Figure 17.5. Four of the primary mechanoreceptors in human skin are shown. Merkel's disks, which are unencapsulated, respond to light touch. Meissner's corpuscles, Ruffini endings, Pacinian corpuscles, and Krause end bulbs are all encapsulated. Meissner's corpuscles respond to touch and low-frequency vibration. Ruffini endings detect stretch, deformation within joints, and warmth. Pacinian corpuscles detect transient pressure and high-frequency vibration. Krause end bulbs detect cold.

Which of the following statements about mechanoreceptors is false?

1. Pacini corpuscles are found in both glabrous and hairy skin.
2. Merkel's disks are abundant on the fingertips and lips.
3. Ruffini endings are encapsulated mechanoreceptors.
4. Meissner's corpuscles extend into the lower dermis.

Meissner's corpuscles, (shown in Figure 17.6) also known as tactile corpuscles, are found in the upper dermis, but they project into the epidermis. They, too, are found primarily in the glabrous skin on the fingertips and eyelids. They respond to fine touch and pressure, but they also respond to low-frequency vibration or flutter. They are rapidly adapting, fluid-filled, encapsulated neurons with small, well-defined borders and are responsive to fine details. Like Merkel's disks, Meissner's corpuscles are not as plentiful in the palms as they are in the fingertips.

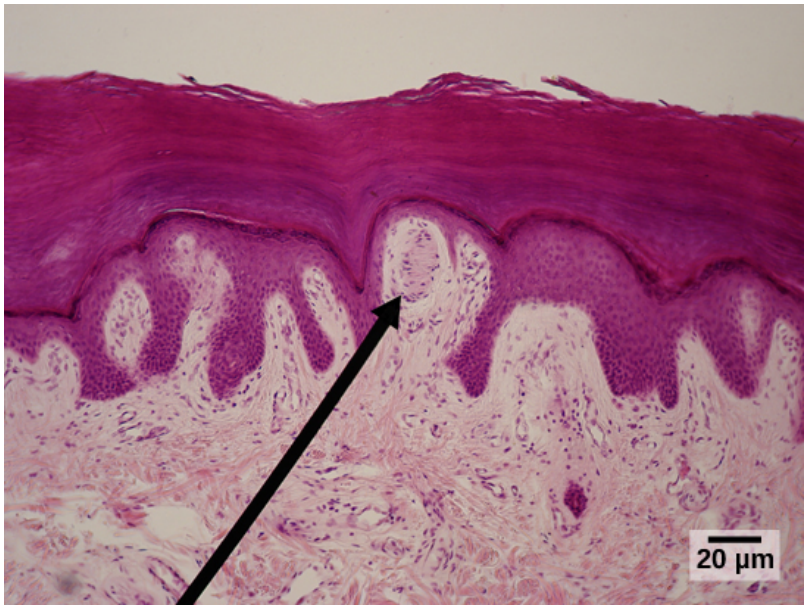


Figure 17.6. Meissner corpuscles in the fingertips, such as the one viewed here using bright field light microscopy, allow for touch discrimination of fine detail. (credit: modification of work by "Wbensmith"/Wikimedia Commons; scale-bar data from Matt Russell)

Deeper in the epidermis, near the base, are **Ruffini endings**, which are also known as bulbous corpuscles. They are found in both glabrous and hairy skin. These are slow-adapting, encapsulated mechanoreceptors that detect skin stretch and deformations within joints, so they provide valuable feedback for gripping objects and controlling finger position and movement. Thus, they also contribute to proprioception and kinesthesia. Ruffini endings also detect warmth. Note that these warmth detectors are situated deeper in the skin than are the cold detectors. It is not surprising, then, that humans detect cold stimuli before they detect warm stimuli.

Pacinian corpuscles (seen in Figure 17.7) are located deep in the dermis of both glabrous and hairy skin and are structurally similar to Meissner's corpuscles; they are found in the bone periosteum, joint capsules, pancreas and other viscera, breast, and genitals. They are rapidly adapting mechanoreceptors that sense deep transient (but not prolonged) pressure and high-frequency vibration. Pacinian receptors detect pressure and vibration by being compressed, stimulating their internal dendrites. There are fewer Pacinian corpuscles and Ruffini endings in skin than there are Merkel's disks and Meissner's corpuscles.

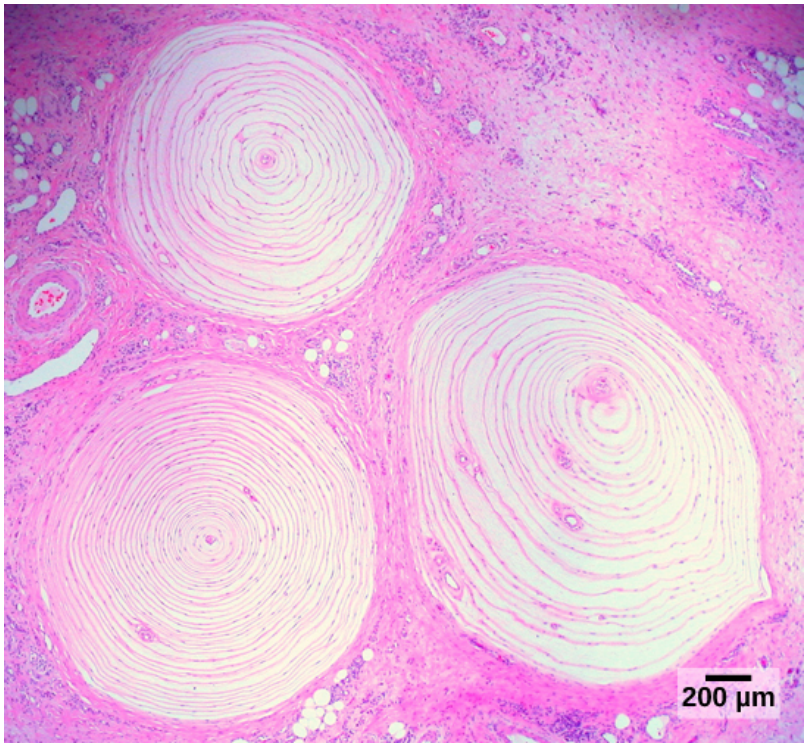


Figure 17.7. Pacinian corpuscles, such as these visualized using bright field light microscopy, detect pressure (touch) and high-frequency vibration. (credit: modification of work by Ed Uthman; scale-bar data from Matt Russell)

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

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

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

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

Integration of Signals from Mechanoreceptors

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

Density of Mechanoreceptors

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

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

Thermoreception

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

Pain

Pain is the name given to **nociception**, which is the neural processing of injurious stimuli in response to tissue damage. Pain is caused by true sources of injury, such as contact with a heat source that causes a thermal burn or contact with a corrosive chemical. But pain also can be caused by harmless stimuli that

mimic the action of damaging stimuli, such as contact with capsaicins, the compounds that cause peppers to taste hot and which are used in self-defense pepper sprays and certain topical medications. Peppers taste “hot” because the protein receptors that bind capsaicin open the same calcium channels that are activated by warm receptors.

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

Concept in Action

View this video that animates the five phases of nociceptive pain.

Summary

Somatosensation includes all sensation received from the skin and mucous membranes, as well as from the limbs and joints. Somatosensation occurs all over the exterior of the body and at some interior locations as well, and a variety of receptor types, embedded in the skin and mucous membranes, play a role.

There are several types of specialized sensory receptors. Rapidly adapting free nerve endings detect nociception, hot and cold, and light touch. Slowly adapting, encapsulated Merkel’s disks are found in fingertips and lips, and respond to light touch. Meissner’s corpuscles, found in glabrous skin, are rapidly adapting, encapsulated receptors that detect touch, low-frequency vibration, and flutter. Ruffini endings are slowly adapting, encapsulated receptors that detect skin stretch, joint activity, and warmth. Hair receptors are rapidly adapting nerve endings wrapped around the base of hair follicles that detect hair movement and skin deflection. Finally, Pacinian corpuscles are encapsulated, rapidly adapting receptors that detect transient pressure and high-frequency vibration.

Exercises

1. _____ are found only in _____ skin, and detect skin deflection.
    1. Meissner's corpuscles: hairy
    2. Merkel's disks: glabrous
    3. hair receptors: hairy
    4. Krause end bulbs: hairy
2. If you were to burn your epidermis, what receptor type would you most likely burn?
    1. free nerve endings
    2. Ruffini endings
    3. Pacinian corpuscle
    4. hair receptors
3. What can be inferred about the relative sizes of the areas of cortex that process signals from skin not densely innervated with sensory receptors and skin that is densely innervated with sensory receptors?

Answers

1. B
2. A
3. The cortical areas serving skin that is densely innervated likely are larger than those serving skin that is less densely innervated.

Glossary

free nerve ending
ending of an afferent neuron that lacks a specialized structure for detection of sensory stimuli; some respond to touch, pain, or temperature

Golgi tendon organ
muscular proprioceptive tension receptor that provides the sensory component of the Golgi tendon reflex

glabrous
describes the non-hairy skin found on palms and fingers, soles of feet, and lips of humans and other primates

Meissner's corpuscle
(also, tactile corpuscle) encapsulated, rapidly-adapting mechanoreceptor in the skin that responds to light touch

Merkel's disc
unencapsulated, slowly-adapting mechanoreceptor in the skin that responds to touch

muscle spindle
proprioceptive stretch receptor that lies within a muscle and that shortens the muscle to an optimal length for efficient contraction

nociception
neural processing of noxious (such as damaging) stimuli

Pacinian corpuscle
encapsulated mechanoreceptor in the skin that responds to deep pressure and vibration

Ruffini ending
(also, bulbous corpuscle) slowly-adapting mechanoreceptor in the skin that responds to skin stretch and joint position

17.3 Taste and Smell

Learning Objectives

By the end of this section, you will be able to:

- Explain in what way smell and taste stimuli differ from other sensory stimuli
- Identify the five primary tastes that can be distinguished by humans
- Explain in anatomical terms why a dog's sense of smell is more acute than a human's

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

Tastes and Odors

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

All odors that we perceive are molecules in the air we breathe. If a substance does not release molecules into the air from its surface, it has no smell. And if a human or other animal does not have a receptor that recognizes a specific molecule, then that molecule has no smell. Humans have about 350 olfactory receptor subtypes that work in various combinations to allow us to sense about 10,000 different odors. Compare that to mice, for example, which have about 1,300 olfactory receptor types, and therefore probably sense more odors. Both odors and tastes involve molecules that stimulate specific chemoreceptors. Although humans commonly distinguish taste as one sense and smell as

another, they work together to create the perception of flavor. A person’s perception of flavor is reduced if he or she has congested nasal passages.

Reception and Transduction

Odorants (odor molecules) enter the nose and dissolve in the olfactory epithelium, the mucosa at the back of the nasal cavity (as illustrated in Figure 17.8). The **olfactory epithelium** is a collection of specialized olfactory receptors in the back of the nasal cavity that spans an area about 5 cm^2 in humans. Recall that sensory cells are neurons. An **olfactory receptor**, which is a dendrite of a specialized neuron, responds when it binds certain molecules inhaled from the environment by sending impulses directly to the olfactory bulb of the brain. Humans have about 12 million olfactory receptors, distributed among hundreds of different receptor types that respond to different odors. Twelve million seems like a large number of receptors, but compare that to other animals: rabbits have about 100 million, most dogs have about 1 billion, and bloodhounds—dogs selectively bred for their sense of smell—have about 4 billion. The overall size of the olfactory epithelium also differs between species, with that of bloodhounds, for example, being many times larger than that of humans.

Olfactory neurons are **bipolar neurons** (neurons with two processes from the cell body). Each neuron has a single dendrite buried in the olfactory epithelium, and extending from this dendrite are 5 to 20 receptor-laden, hair-like cilia that trap odorant molecules. The sensory receptors on the cilia are proteins, and it is the variations in their amino acid chains that make the receptors sensitive to different odorants. Each olfactory sensory neuron has only one type of receptor on its cilia, and the receptors are specialized to detect specific odorants, so the bipolar neurons themselves are specialized. When an odorant binds with a receptor that recognizes it, the sensory neuron associated with the receptor is stimulated. Olfactory stimulation is the only sensory information that directly reaches the cerebral cortex, whereas other sensations are relayed through the thalamus.

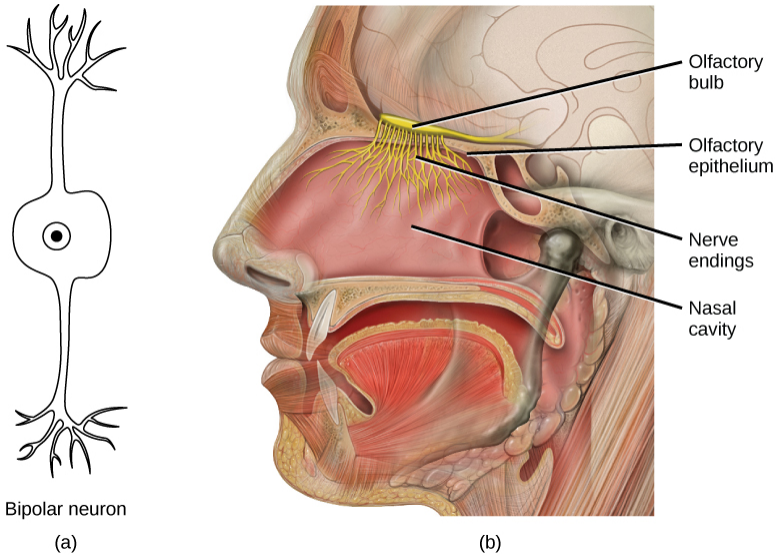


Figure 17.8. In the human olfactory system, (a) bipolar olfactory neurons extend from (b) the olfactory epithelium, where olfactory receptors are located, to the olfactory bulb. (credit: modification of work by Patrick J. Lynch, medical illustrator; C. Carl Jaffe, MD, cardiologist)

Pheromones

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

The vomeronasal organ (VNO, or Jacobson's organ) is a tubular, fluid-filled, olfactory organ present in many vertebrate animals that sits adjacent to the nasal cavity. It is very sensitive to pheromones and is connected to the nasal cavity by a duct. When molecules dissolve in the mucosa of the nasal cavity, they then enter the VNO where the pheromone molecules among them bind with specialized pheromone receptors. Upon exposure to pheromones from their own species or others, many animals, including cats, may display the flehmen response (shown in Figure 17.9), a curling of the upper lip that helps pheromone molecules enter the VNO.

Pheromonal signals are sent, not to the main olfactory bulb, but to a different neural structure that projects directly to the amygdala (recall that the amygdala is a brain center important in emotional reactions, such as fear). The pheromonal signal then continues to areas of the hypothalamus that are key to reproductive physiology and behavior. While some scientists assert that the VNO is apparently functionally vestigial in humans, even though there is a similar structure located near human nasal cavities, others are researching it as a possible functional system that may, for example, contribute to synchronization of menstrual cycles in women living in close proximity.

Figure 17.9.
The flehmen response in this tiger results in the curling of the upper lip and helps airborne pheromone molecules enter the vomeronasal organ. (credit: modification of work by "chadh"/Flickr)

Taste

Detecting a taste (gustation) is fairly similar to detecting an odor (olfaction), given that both taste and smell rely on chemical receptors being stimulated by certain molecules. The primary organ of taste is the taste bud. A *taste bud* is a cluster of gustatory receptors (taste cells) that are located within the bumps on the tongue called *papillae* (singular: papilla) (illustrated in Figure 17.10). There are several structurally distinct papillae. Filiform papillae, which are located across the tongue, are tactile, providing friction that helps the tongue move substances, and contain no taste cells. In contrast, fungiform papillae, which are located mainly on the anterior two-thirds of the tongue, each contain one to eight taste buds and also have receptors for pressure and temperature. The large circumvallate papillae contain up to 100 taste buds and form a V near the posterior margin of the tongue.

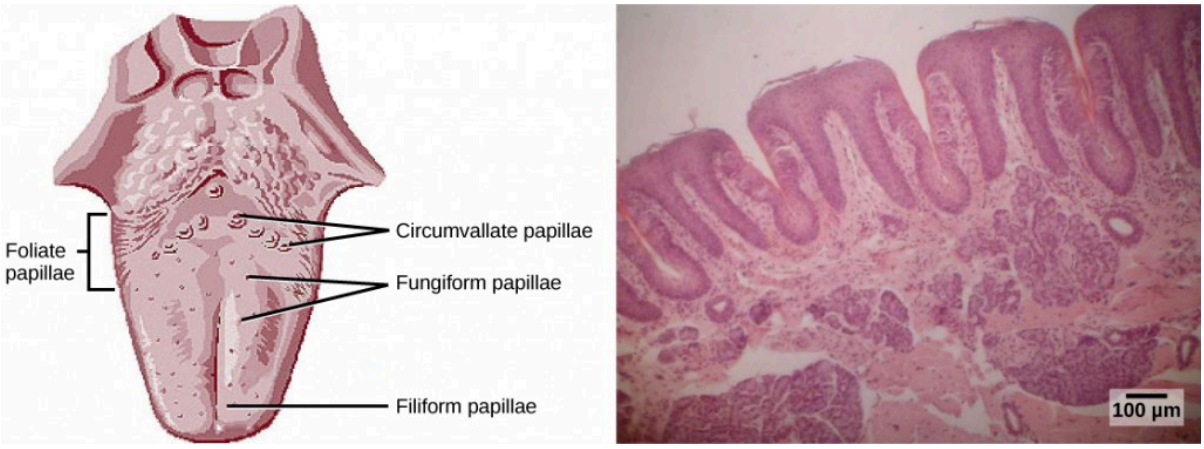


Figure 17.10. (a) Foliate, circumvallate, and fungiform papillae are located on different regions of the tongue. (b) Foliate papillae are prominent protrusions on this light micrograph. (credit a: modification of work by NCI; scale-bar data from Matt Russell)

In addition to those two types of chemically and mechanically sensitive papillae are foliate papillae—leaf-like papillae located in parallel folds along the edges and toward the back of the tongue, as seen in the Figure 17.10 micrograph. Foliate papillae contain about 1,300 taste buds within their folds. Finally, there are circumvallate papillae, which are wall-like papillae in the shape of an inverted "V" at the back of the tongue. Each of these papillae is surrounded by a groove and contains about 250 taste buds.

Each taste bud's taste cells are replaced every 10 to 14 days. These are elongated cells with hair-like processes called microvilli at the tips that extend into the taste bud pore (illustrate in Figure 17.11). 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.

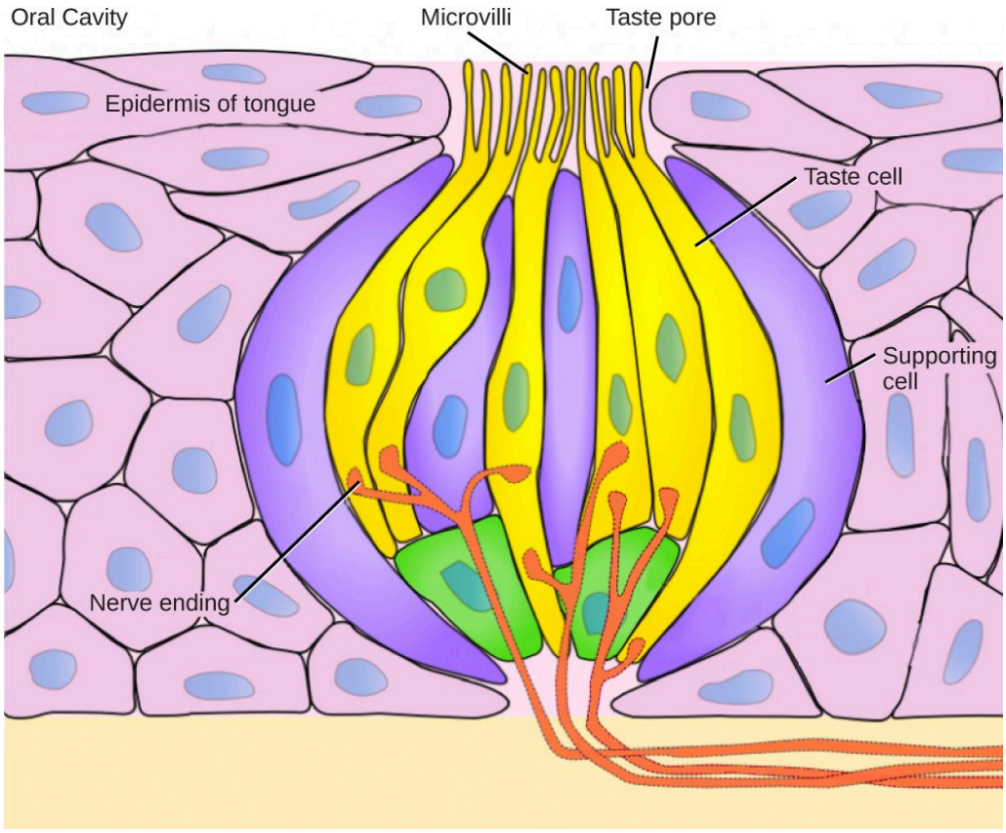


Figure 17.11. Pores in the tongue allow tastants to enter taste pores in the tongue. (credit: modification of work by Vincenzo Rizzo)

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

Both tasting abilities and sense of smell change with age. In humans, the senses decline dramatically by age 50 and continue to decline. A child may find a food to be too spicy, whereas an elderly person may find the same food to be bland and unappetizing.

Concept in Action

View this animation that shows how the sense of taste works.

Smell and Taste in the Brain

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

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

Summary

There are five primary tastes in humans: sweet, sour, bitter, salty, and umami. Each taste has its own receptor type that responds only to that taste. Tastants enter the body and are dissolved in saliva. Taste cells are located within taste buds, which are found on three of the four types of papillae in the mouth.

Regarding olfaction, there are many thousands of odorants, but humans detect only about 10,000. Like taste receptors, olfactory receptors are each responsive to only one odorant. Odorants dissolve in nasal mucosa, where they excite their corresponding olfactory sensory cells. When these cells detect an odorant, they send their signals to the main olfactory bulb and then to other locations in the brain, including the olfactory cortex.

Exercises

1. Which of the following has the fewest taste receptors?
    1. fungiform papillae
    2. circumvallate papillae
    3. foliate papillae
    4. filiform papillae
2. How many different taste molecules do taste cells each detect?
    1. one
    2. five
    3. ten
    4. It depends on the spot on the tongue
3. Salty foods activate the taste cells by _____.
    1. exciting the taste cell directly
    2. causing hydrogen ions to enter the cell
    3. causing sodium channels to close
    4. binding directly to the receptors
4. All sensory signals except _____ travel to the _____ in the brain before the cerebral cortex.
    1. vision; thalamus
    2. olfaction; thalmus
    3. vision; cranial nerves
    4. olfaction; cranial nerves
5. From the perspective of the recipient of the signal, in what ways do pheromones differ from other odorants?
6. What might be the effect on an animal of not being able to perceive taste?

Answers

1. D
2. A
3. A
4. B
5. Pheromones may not be consciously perceived, and pheromones can have direct physiological and behavioral effects on their recipients.
6. The animal might not be able to recognize the differences in food sources and thus might not be able to discriminate between spoiled food and safe food or between foods that contain necessary

nutrients, such as proteins, and foods that do not.

Exercises

1. Which of the following has the fewest taste receptors?

A) fungiform papillae
B) circumvallate papillae
C) foliate papillae
D) filiform papillae

Answer: D

2. How many different taste molecules do taste cells each detect?

A) one
B) five
C) ten
D) It depends on the spot on the tongue
Answer: A

3. Salty foods activate the taste cells by _____.

A) exciting the taste cell directly
B) causing hydrogen ions to enter the cell
C) causing sodium channels to close
D) binding directly to the receptors

Answer: A

4. All sensory signals except _____ travel to the _____ in the brain before the cerebral cortex.

A) vision; thalamus
B) olfaction; thalamus
C) vision; cranial nerves
D) olfaction; cranial nerves
Answer: B

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

Pheromones may not be consciously perceived, and pheromones can have direct physiological and behavioral effects on their recipients.

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

The animal might not be able to recognize the differences in food sources and thus might not be able to discriminate between spoiled food and safe food or between foods that contain necessary nutrients, such as proteins, and foods that do not.

Glossary

bipolar neuron
neuron with two processes from the cell body, typically in opposite directions

glomerulus
in the olfactory bulb, one of the two neural clusters that receives signals from one type of olfactory receptor

gustation
sense of taste

odorant
airborne molecule that stimulates an olfactory receptor

olfaction
sense of smell

olfactory epithelium
specialized tissue in the nasal cavity where olfactory receptors are located

olfactory receptor
dendrite of a specialized neuron

pheromone
substance released by an animal that can affect the physiology or behavior of other animals

tastant
food molecule that stimulates gustatory receptors

umami
one of the five basic tastes, which is described as “savory” and which may be largely the taste of L-glutamate

17.4 Hearing and Vestibular Sensation

Learning Objectives

By the end of this section, you will be able to:

- Describe the relationship of amplitude and frequency of a sound wave to attributes of sound
- Trace the path of sound through the auditory system to the site of transduction of sound
- Identify the structures of the vestibular system that respond to gravity

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

Sound

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

As is true for all waves, there are four main characteristics of a sound wave: frequency, wavelength, period, and amplitude. Frequency is the number of waves per unit of time, and in sound is heard as pitch. High-frequency (≥15.000Hz) sounds are higher-pitched (short wavelength) than low-frequency (long wavelengths; ≤100Hz) sounds. Frequency is measured in cycles per second, and for sound, the most commonly used unit is hertz (Hz), or cycles per second. Most humans can perceive sounds with frequencies between 30 and 20,000 Hz. Women are typically better at hearing high frequencies, but everyone's ability to hear high frequencies decreases with age. Dogs detect up to about 40,000 Hz; cats, 60,000 Hz; bats, 100,000 Hz; and dolphins 150,000 Hz, and American shad (**Alosa sapidissima**), a fish, can hear 180,000 Hz. Those frequencies above the human range are called **ultrasound**.

Amplitude, or the dimension of a wave from peak to trough, in sound is heard as volume and is illustrated in Figure 17.12. The sound waves of louder sounds have greater amplitude than those of softer sounds. For sound, volume is measured in decibels (dB). The softest sound that a human can hear is the zero point. Humans speak normally at 60 decibels.

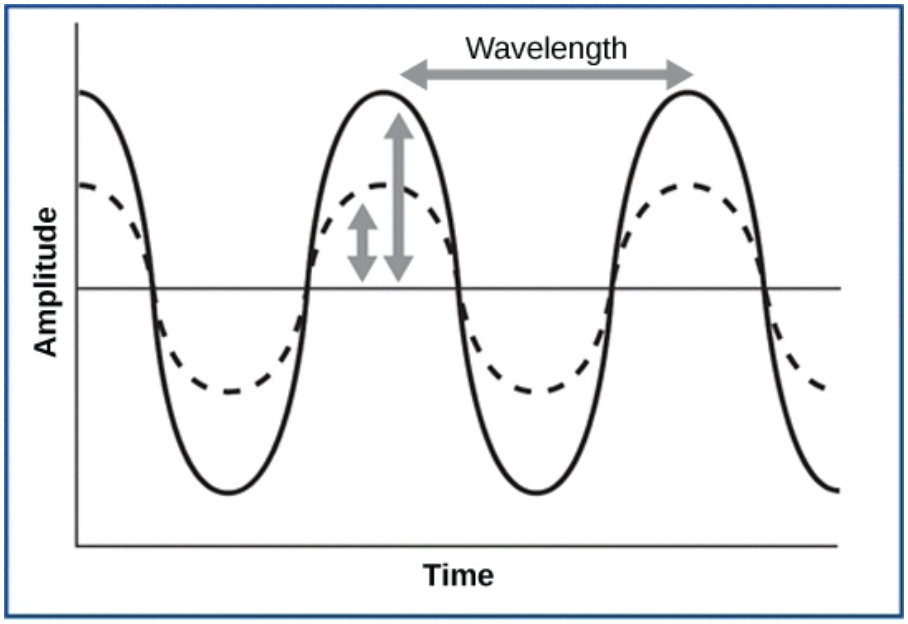


Figure 17.12. For sound waves, wavelength corresponds to pitch. Amplitude of the wave corresponds to volume. The sound wave shown with a dashed line is softer in volume than the sound wave shown with a solid line. (credit: NIH)

Reception of Sound

In mammals, sound waves are collected by the external, cartilaginous part of the ear called the **pinna**, then travel through the auditory canal and cause vibration of the thin diaphragm called the **tympanum** or ear drum, the innermost part of the **outer ear** (illustrated in Figure 17.13). Interior to the tympanum is the **middle ear**. The middle ear holds three small bones called the **ossicles,** which transfer energy from the moving tympanum to the inner ear. The three ossicles are the **malleus** (also known as the hammer), the **incus** (the anvil), and **stapes** (the stirrup). The aptly named stapes looks very much like a stirrup. The three ossicles are unique to mammals, and each plays a role in hearing. The malleus attaches at three points to the interior surface of the tympanic membrane. The incus attaches the malleus to the stapes. In humans, the stapes is not long enough to reach the tympanum. If we did not have the malleus and the incus, then the vibrations of the tympanum would never reach the inner ear. These bones also function to collect force and amplify sounds. The ear ossicles are homologous to bones in a fish mouth: the bones that support gills in fish are thought to be adapted for use in the vertebrate ear over evolutionary time. Many animals (frogs, reptiles, and birds, for example) use the stapes of the middle ear to transmit vibrations to the middle ear.

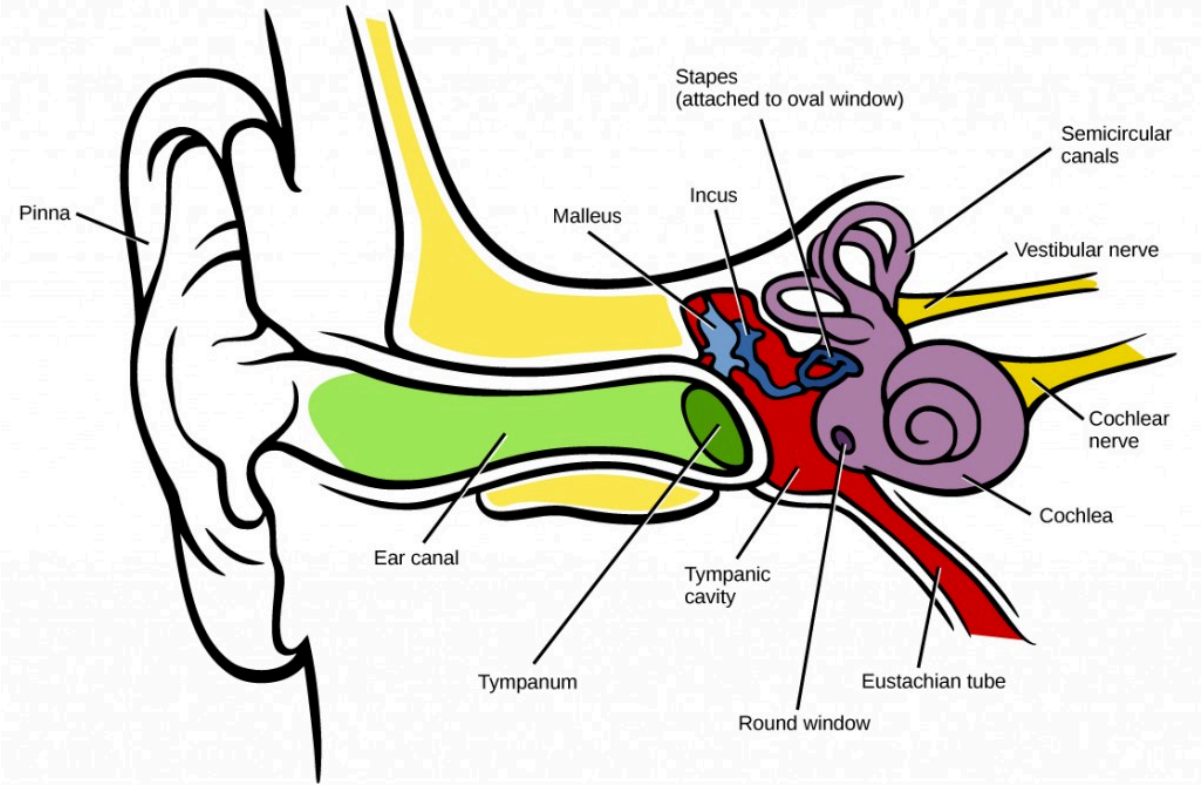


Figure 17.13. Sound travels through the outer ear to the middle ear, which is bounded on its exterior by the tympanic membrane. The middle ear contains three bones called ossicles that transfer the sound wave to the oval window, the exterior boundary of the inner ear. The organ of Corti, which is the organ of sound transduction, lies inside the cochlea. (credit: modification of work by Lars Chittka, Axel Brockmann)

Transduction of Sound

Vibrating objects, such as vocal cords, create sound waves or pressure waves in the air. When these pressure waves reach the ear, the ear transduces this mechanical stimulus (pressure wave) into a nerve impulse (electrical signal) that the brain perceives as sound. The pressure waves strike the tympanum, causing it to vibrate. The mechanical energy from the moving tympanum transmits the vibrations to the three bones of the middle ear. The stapes transmits the vibrations to a thin diaphragm called the **oval window**, which is the outermost structure of the **inner ear**. The structures of the inner ear are found in the **labyrinth**, a bony, hollow structure that is the most interior portion of the ear. Here, the energy from the sound wave is transferred from the stapes through the flexible oval window and to the fluid of the cochlea. The vibrations of the oval window create pressure waves in the fluid (perilymph) inside the cochlea. The **cochlea** is a whorled structure, like the shell of a snail, and it contains receptors for transduction of the mechanical wave into an electrical signal (as illustrated in Figure 17.14). Inside the cochlea, the **basilar membrane** is a mechanical analyzer that runs the length of the cochlea, curling toward the cochlea's center.

The mechanical properties of the basilar membrane change along its length, such that it is thicker, tauter, and narrower at the outside of the whorl (where the cochlea is largest), and thinner, floppier, and broader toward the apex, or center, of the whorl (where the cochlea is smallest). Different regions of the basilar membrane vibrate according to the frequency of the sound wave conducted through the fluid in the cochlea. For these reasons, the fluid-filled cochlea detects different wave frequencies (pitches) at

different regions of the membrane. When the sound waves in the cochlear fluid contact the basilar membrane, it flexes back and forth in a wave-like fashion. Above the basilar membrane is the **tectorial membrane**.

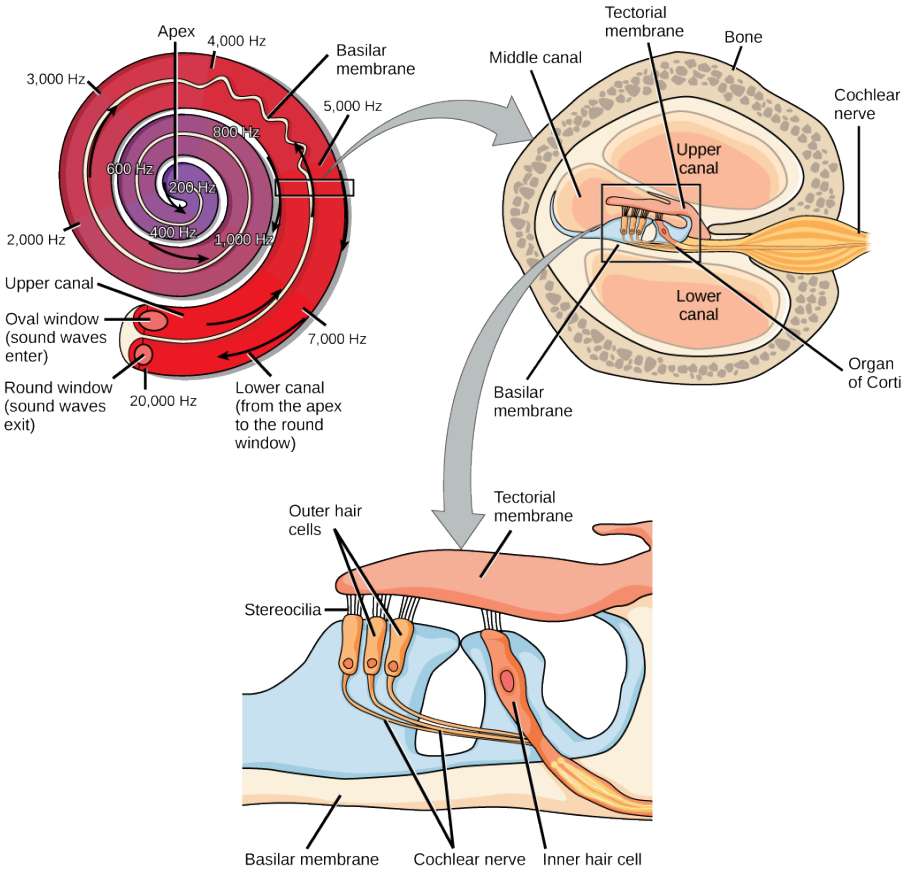


Figure 17.14. In the human ear, sound waves cause the stapes to press against the oval window. Vibrations travel up the fluid-filled interior of the cochlea. The basilar membrane that lines the cochlea gets continuously thinner toward the apex of the cochlea. Different thicknesses of membrane vibrate in response to different frequencies of sound. Sound waves then exit through the round window. In the cross section of the cochlea (top right figure), note that in addition to the upper canal and lower canal, the cochlea also has a middle canal. The organ of Corti (bottom image) is the site of sound transduction. Movement of stereocilia on hair cells results in an action potential that travels along the auditory nerve.

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

1. Hearing loss resulting from absence or loss of hair cells in the organ of Corti.
2. Hearing loss resulting from an abnormal auditory nerve.
3. Hearing loss resulting from fracture of the cochlea.

4. Hearing loss resulting from damage to bones of the middle ear.

The site of transduction is in the **organ of Corti** (spiral organ). It is composed of hair cells held in place above the basilar membrane like flowers projecting up from soil, with their exposed short, hair-like **stereocilia** contacting or embedded in the tectorial membrane above them. The inner hair cells are the primary auditory receptors and exist in a single row, numbering approximately 3,500. The stereocilia from inner hair cells extend into small dimples on the tectorial membrane's lower surface. The outer hair cells are arranged in three or four rows. They number approximately 12,000, and they function to fine tune incoming sound waves. The longer stereocilia that project from the outer hair cells actually attach to the tectorial membrane. All of the stereocilia are mechanoreceptors, and when bent by vibrations they respond by opening a gated ion channel (refer to Figure 17.2). As a result, the hair cell membrane is depolarized, and a signal is transmitted to the chochlear nerve. Intensity (volume) of sound is determined by how many hair cells at a particular location are stimulated.

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

When sound waves produce fluid waves inside the cochlea, the basilar membrane flexes, bending the stereocilia that attach to the tectorial membrane. Their bending results in action potentials in the hair cells, and auditory information travels along the neural endings of the bipolar neurons of the hair cells (collectively, the auditory nerve) to the brain. When the hairs bend, they release an excitatory neurotransmitter at a synapse with a sensory neuron, which then conducts action potentials to the central nervous system. The cochlear branch of the vestibulocochlear cranial nerve sends information on hearing. The auditory system is very refined, and there is some modulation or "sharpening" built in. The brain can send signals back to the cochlea, resulting in a change of length in the outer hair cells, sharpening or dampening the hair cells' response to certain frequencies.

Concept in Action

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

Higher Processing

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

Vestibular Information

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

The vestibular system has some similarities with the auditory system. It utilizes hair cells just like the auditory system, but it excites them in different ways. There are five vestibular receptor organs in the inner ear: the utricle, the saccule, and three semicircular canals. Together, they make up what's known as the vestibular labyrinth that is shown in Figure 17.15. 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.

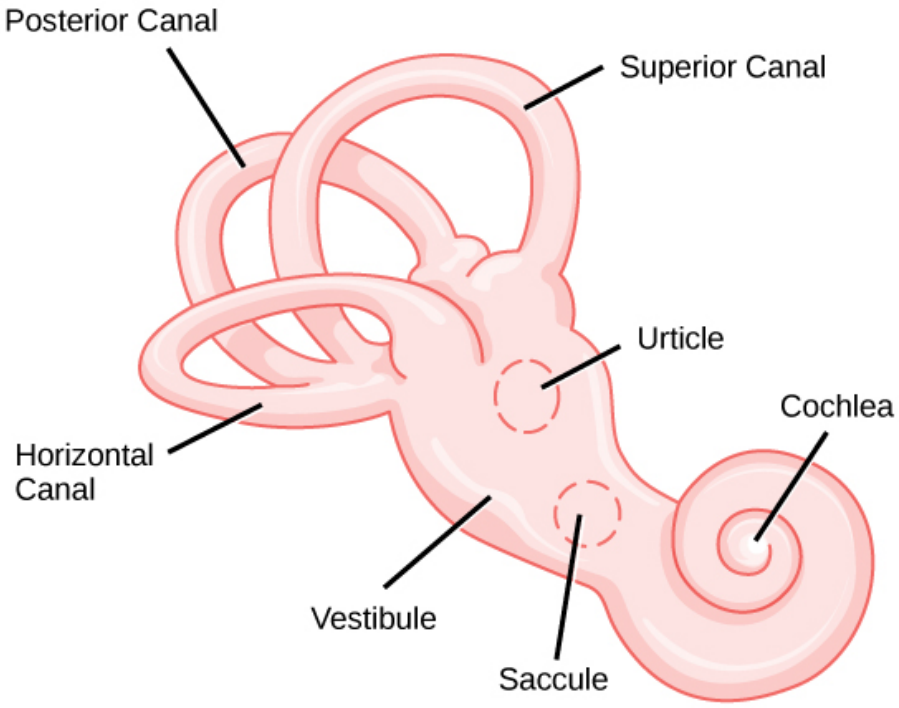


Figure 17.15. The structure of the vestibular labyrinth is shown. (credit: modification of work by NIH)

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 17.15. 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.

Higher Processing

Hair cells from the utricle, saccule, and semicircular canals also communicate through bipolar neurons to the cochlear nucleus in the medulla. Cochlear neurons send descending projections to the spinal cord and ascending projections to the pons, thalamus, and cerebellum. Connections to the cerebellum are important for coordinated movements. There are also projections to the temporal cortex, which account for feelings of dizziness; projections to autonomic nervous system areas in the brainstem, which account for motion sickness; and projections to the primary somatosensory cortex, which monitors subjective measurements of the external world and self-movement. People with lesions in the vestibular

area of the somatosensory cortex see vertical objects in the world as being tilted. Finally, the vestibular signals project to certain optic muscles to coordinate eye and head movements.

Concept in Action

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

Summary

Audition is important for territory defense, predation, predator defense, and communal exchanges. The vestibular system, which is not auditory, detects linear acceleration and angular acceleration and deceleration. Both the auditory system and vestibular system use hair cells as their receptors.

Auditory stimuli are sound waves. The sound wave energy reaches the outer ear (pinna, canal, tympanum), and vibrations of the tympanum send the energy to the middle ear. The middle ear bones shift and the stapes transfers mechanical energy to the oval window of the fluid-filled inner ear cochlea. Once in the cochlea, the energy causes the basilar membrane to flex, thereby bending the stereocilia on receptor hair cells. This activates the receptors, which send their auditory neural signals to the brain.

The vestibular system has five parts that work together to provide the sense of direction, thus helping to maintain balance. The utricle and saccule measure head orientation: their calcium carbonate crystals shift when the head is tilted, thereby activating hair cells. The semicircular canals work similarly, such that when the head is turned, the fluid in the canals bends stereocilia on hair cells. The vestibular hair cells also send signals to the thalamus and to somatosensory cortex, but also to the cerebellum, the structure above the brainstem that plays a large role in timing and coordination of movement.

Exercises

1. Which of the following types of hearing loss would not be restored by a cochlear implant?
    1. Hearing loss resulting from absence or loss of hair cells in the organ of Corti.
    2. Hearing loss resulting from an abnormal auditory nerve.
    3. Hearing loss resulting from fracture of the cochlea.
    4. Hearing loss resulting from damage to bones of the middle ear.
2. In sound, pitch is measured in _____, and volume is measured in _____.

1. nanometers (nm); decibels (dB)
2. decibels (dB); nanometers (nm)
3. decibels (dB); hertz (Hz)
4. hertz (Hz); decibels (dB)

3. Auditory hair cells are indirectly anchored to the _____.
    1. basilar membrane
    2. oval window
    3. tectorial membrane
    4. ossicles
4. Which of the following are found both in the auditory system and the vestibular system?
    1. basilar membrane
    2. hair cells
    3. semicircular canals
    4. ossicles
5. How would a rise in altitude likely affect the speed of a sound transmitted through air?
6. How might being in a place with less gravity than Earth has (such as Earth's moon) affect vestibular sensation, and why?

Answers

1. B
2. D
3. A
4. B
5. The sound would slow down, because it is transmitted through the particles (gas) and there are fewer particles (lower density) at higher altitudes.
6. Because vestibular sensation relies on gravity's effects on tiny crystals in the inner ear, a situation of reduced gravity would likely impair vestibular sensation.

Glossary

audition
sense of hearing

basilar membrane
stiff structure in the cochlea that indirectly anchors auditory receptors

cochlea
whorled structure that contains receptors for transduction of the mechanical wave into an electrical signal
s a specialized structure for detection of sensory stimuli; some respond to touch, pain, or temperature

incus
(also, anvil) second of the three bones of the middle ear

inner ear
innermost part of the ear; consists of the cochlea and the vestibular system

labyrinth
bony, hollow structure that is the most internal part of the ear; contains the sites of transduction of auditory and vestibular information

malleus
(also, hammer) first of the three bones of the middle ear

middle ear
part of the hearing apparatus that functions to transfer energy from the tympanum to the oval window of the inner ear

organ of Corti
in the basilar membrane, the site of the transduction of sound, a mechanical wave, to a neural signal

outer ear
part of the ear that consists of the pinna, ear canal, and tympanum and which conducts sound waves into the middle ear

oval window
thin diaphragm between the middle and inner ears that receives sound waves from contact with the stapes bone of the middle ear

pinna
cartilaginous outer ear

semicircular canal
one of three half-circular, fluid-filled tubes in the vestibular labyrinth that monitors angular acceleration and deceleration

stapes
(also, stirrup) third of the three bones of the middle ear

tectorial membrane
cochlear structure that lies above the hair cells and participates in the transduction of sound at the hair cells

tympanum
(also, tympanic membrane or ear drum) thin diaphragm between the outer and middle ears

ultrasound
sound frequencies above the human detectable ceiling of approximately 20,000 Hz

17.5 Vision

Learning Objectives

By the end of this section, you will be able to:

- Explain how electromagnetic waves differs from sound waves
- Trace the path of light through the eye to the point of the optic nerve
- Explain tonic activity as it is manifested in photoreceptors in the retina

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 17.16). 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).

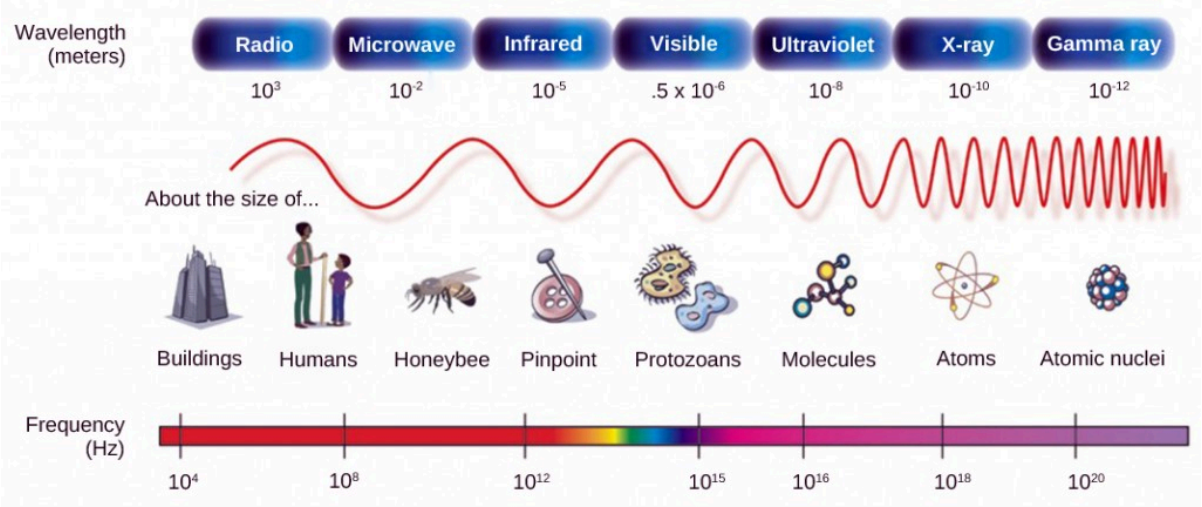


Figure 17.16. In the electromagnetic spectrum, visible light lies between 380 nm and 740 nm. (credit: modification of work by NASA)

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 17.17) 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 17.17**b**). 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.

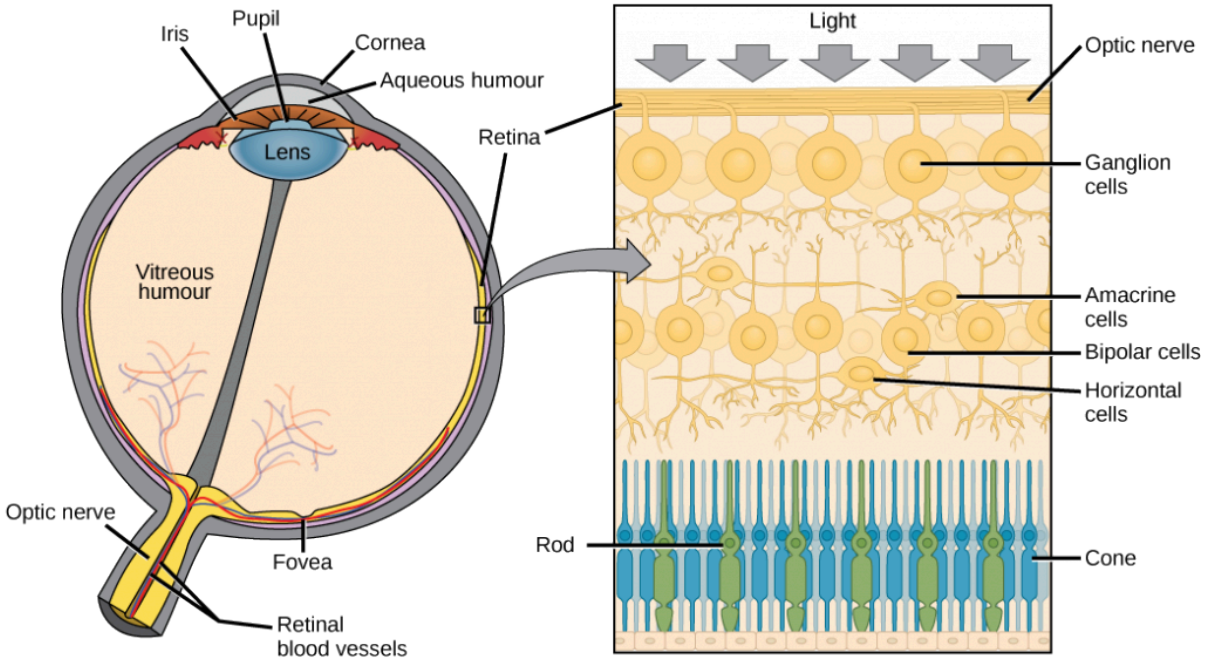


Figure 17.17. (a) The human eye is shown in cross section. (b) A blowup shows the layers of the retina.

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

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

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 17.18. 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.

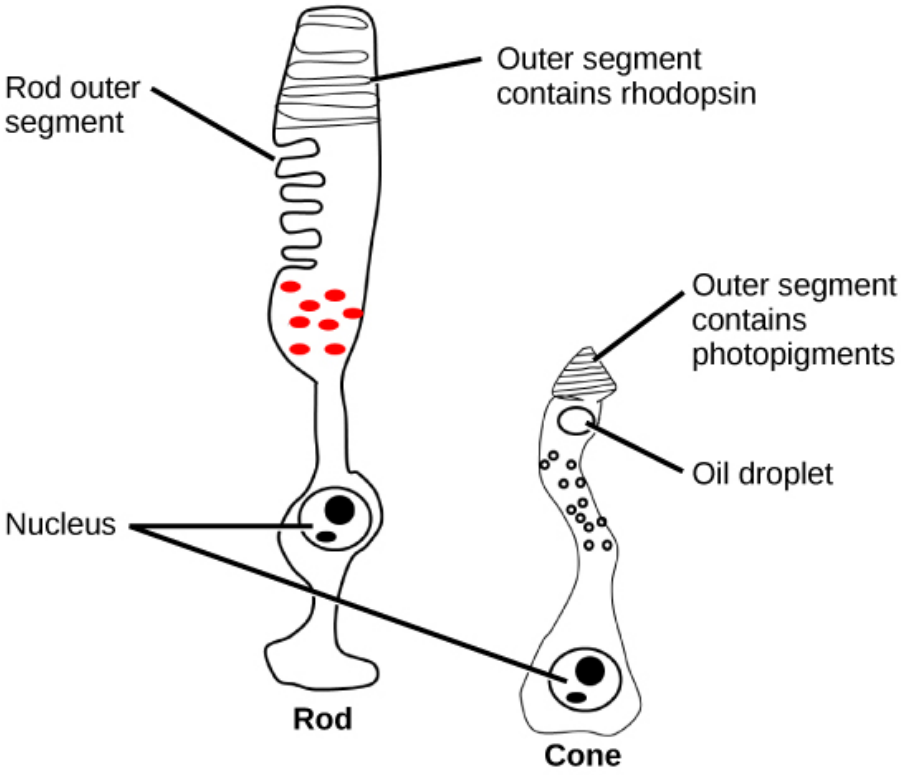


Figure 17.18. Rods and cones are photoreceptors in the retina. Rods respond in low light and can detect only shades of gray. Cones respond in intense light and are responsible for color vision. (credit: modification of work by Piotr Sliwa)

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.

Concept in Action

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

Transduction of Light

The rods and cones are the site of transduction of light to a neural signal. Both rods and cones contain photopigments. In vertebrates, the main photopigment, **rhodopsin**, has two main parts Figure 17.19): an opsin, which is a membrane protein (in the form of a cluster of α-helices that span the membrane), and retinal—a molecule that absorbs light. When light hits a photoreceptor, it causes a shape change in the retinal, altering its structure from a bent (*cis*) form of the molecule to its linear (*trans*) isomer. This isomerization of retinal activates the rhodopsin, starting a cascade of events that ends with the closing of Na^+ channels in the membrane of the photoreceptor. Thus, unlike most other sensory neurons (which become depolarized by exposure to a stimulus) visual receptors become hyperpolarized and thus driven away from threshold (Figure 17.20).

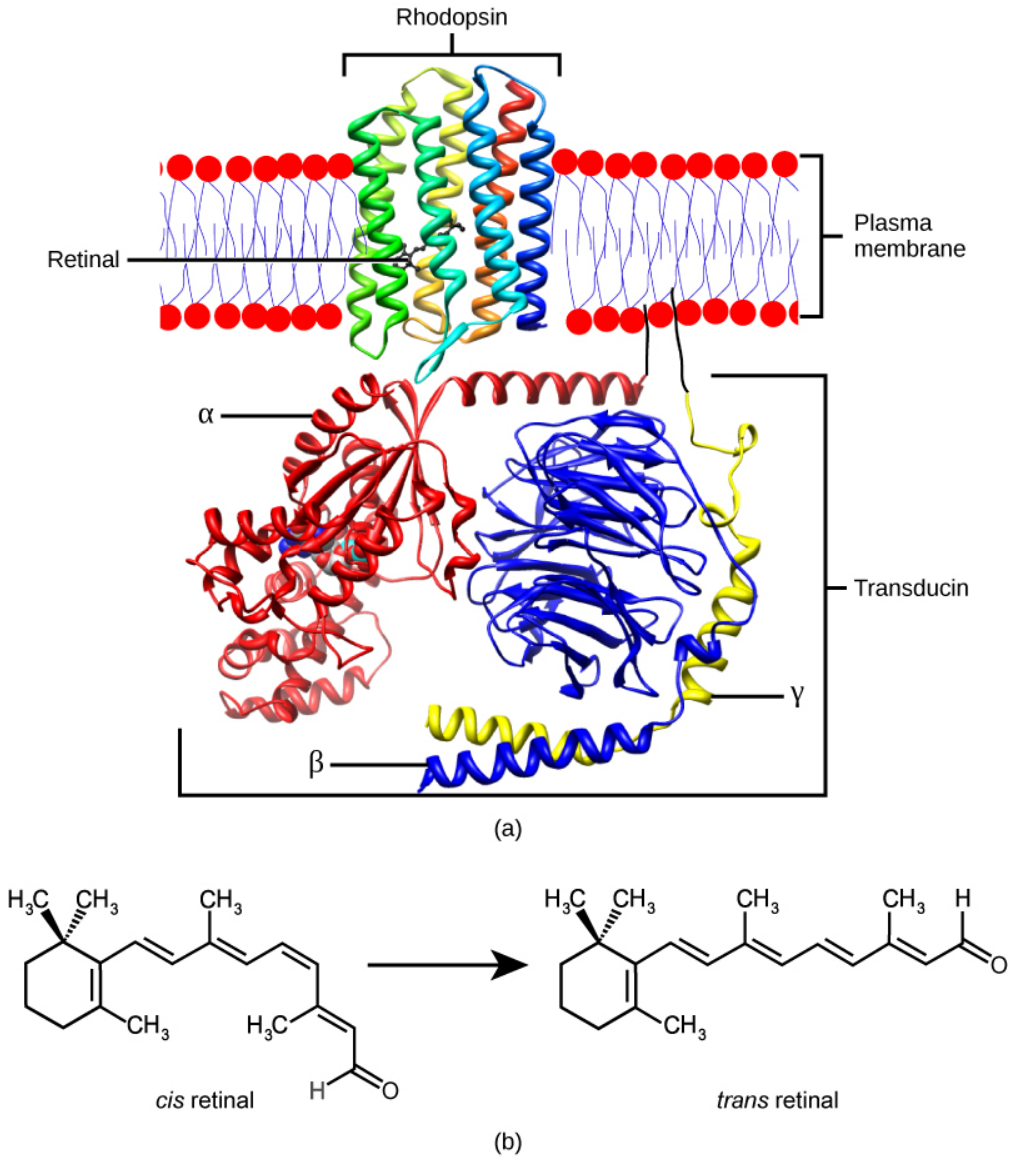


Figure 17.19. (a) Rhodopsin, the photoreceptor in vertebrates, has two parts: the trans-membrane protein opsin, and retinal. When light strikes retinal, it changes shape from (b) a cis to a trans form. The signal is passed to a G-protein called transducin, triggering a series of downstream events.

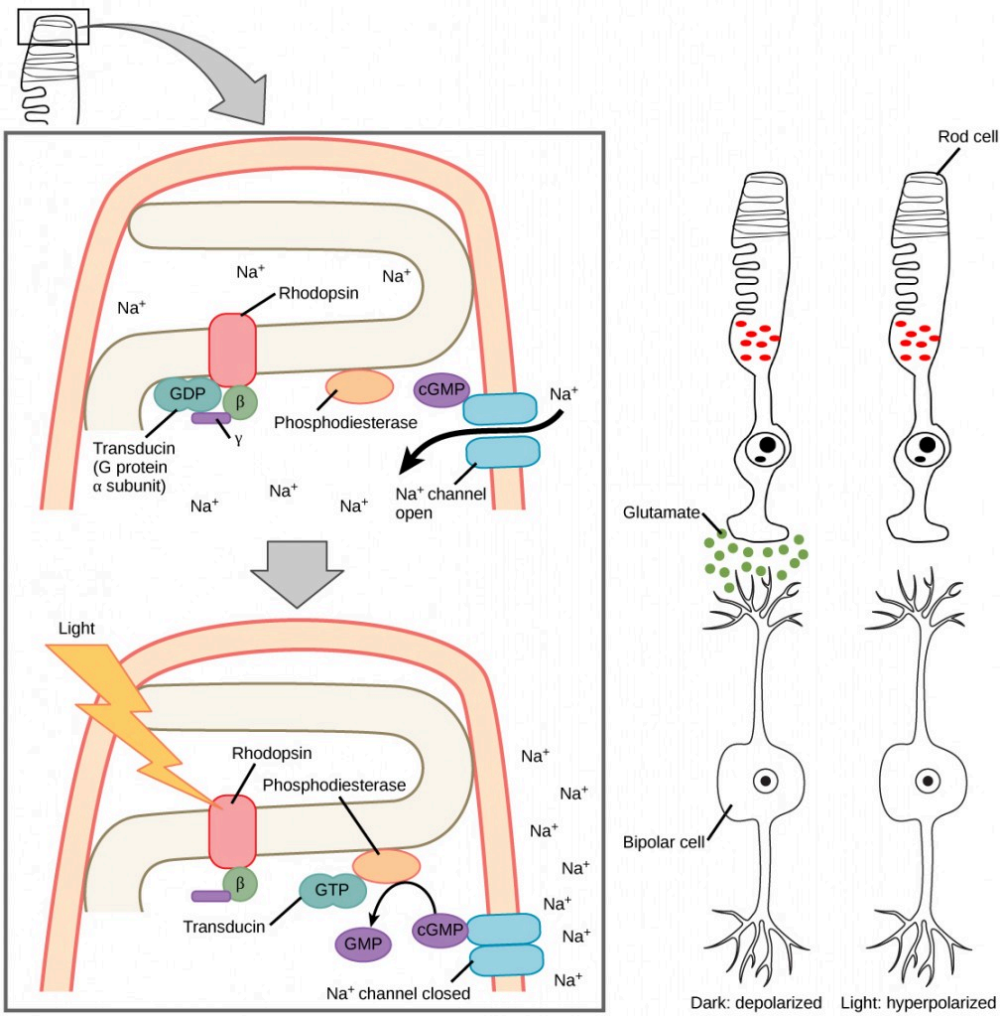


Figure 17.20.
When light strikes rhodopsin, the G-protein transducin is activated, which in turn activates phosphodiesterase. Phosphodiesterase converts cGMP to GMP, thereby closing sodium channels. As a result, the membrane becomes hyperpolarized. The hyperpolarized membrane does not release glutamate to the bipolar cell.

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 17.21. 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.

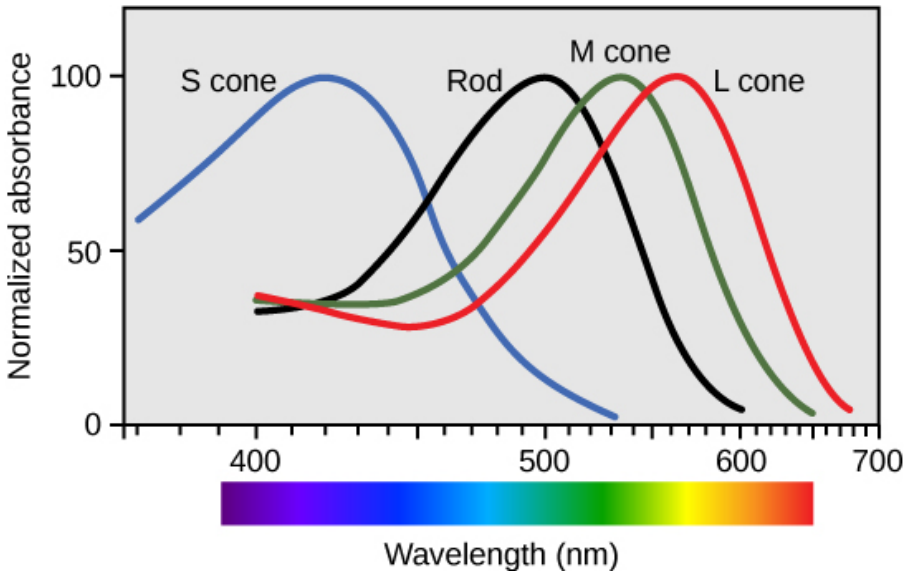


Figure 17.21.
Human rod cells and the different types of cone cells each have an optimal wavelength. However, there is considerable overlap in the wavelengths of light detected.

Retinal Processing

Visual signals leave the cones and rods, travel to the bipolar cells, and then to ganglion cells. A large degree of processing of visual information occurs in the retina itself, before visual information is sent to the brain.

Photoreceptors in the retina continuously undergo **tonic activity**. That is, they are always slightly active even when not stimulated by light. In neurons that exhibit tonic activity, the absence of stimuli maintains a firing rate at a baseline; while some stimuli increase firing rate from the baseline, and other stimuli decrease firing rate. In the absence of light, the bipolar neurons that connect rods and cones to ganglion cells are continuously and actively inhibited by the rods and cones. Exposure of the retina to light hyperpolarizes the rods and cones and removes their inhibition of bipolar cells. The now active bipolar cells in turn stimulate the ganglion cells, which send action potentials along their axons (which leave the eye as the optic nerve). Thus, the visual system relies on change in retinal activity, rather than the absence or presence of activity, to encode visual signals for the brain. Sometimes horizontal cells carry signals from one rod or cone to other photoreceptors and to several bipolar cells. When a rod or cone stimulates a horizontal cell, the horizontal cell inhibits more distant photoreceptors and bipolar cells, creating lateral inhibition. This inhibition sharpens edges and enhances contrast in the images by making regions receiving light appear lighter and dark surroundings appear darker. Amacrine cells can distribute information from one bipolar cell to many ganglion cells.

You can demonstrate this using an easy demonstration to “trick” your retina and brain about the colors you are observing in your visual field. Look fixedly at Figure 17.22 for about 45 seconds. Then quickly shift your gaze to a sheet of blank white paper or a white wall. You should see an afterimage of the

Norwegian flag in its correct colors. At this point, close your eyes for a moment, then reopen them, looking again at the white paper or wall; the afterimage of the flag should continue to appear as red, white, and blue. What causes this? According to an explanation called opponent process theory, as you gazed fixedly at the green, black, and yellow flag, your retinal ganglion cells that respond positively to green, black, and yellow increased their firing dramatically. When you shifted your gaze to the neutral white ground, these ganglion cells abruptly decreased their activity and the brain interpreted this abrupt downshift as if the ganglion cells were responding now to their "opponent" colors: red, white, and blue, respectively, in the visual field. Once the ganglion cells return to their baseline activity state, the false perception of color will disappear.

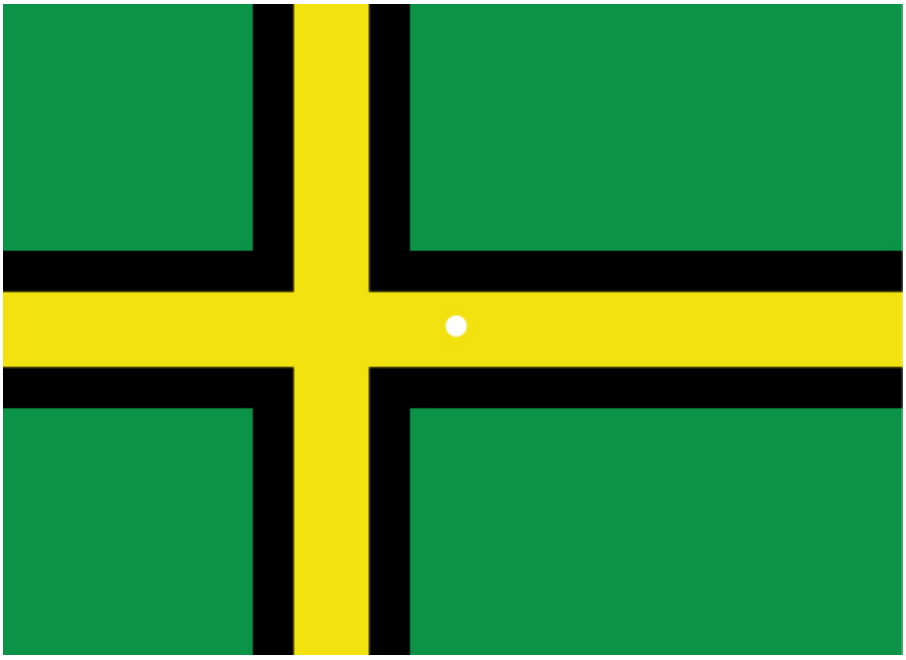


Figure 17.22. View this flag to understand how retinal processing works. Stare at the center of the flag (indicated by the white dot) for 45 seconds, and then quickly look at a white background, noticing how colors appear.

Higher Processing

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

Once in the brain, visual information is processed in several places, and its routes reflect the complexity and importance of visual information to humans and other animals. One route takes the signals to the thalamus, which serves as the routing station for all incoming sensory impulses except olfaction. In the thalamus, the magnocellular and parvocellular distinctions remain intact, and there are different layers of the thalamus dedicated to each. When visual signals leave the thalamus, they travel

to the primary visual cortex at the rear of the brain. From the visual cortex, the visual signals travel in two directions. One stream that projects to the parietal lobe, in the side of the brain, carries magnocellular (“where”) information. A second stream projects to the temporal lobe and carries both magnocellular (“where”) and parvocellular (“what”) information.

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

Concept in Action

View this

interactive presentation to review what you have learned about how vision functions.

Summary

Vision is the only photo responsive sense. Visible light travels in waves and is a very small slice of the electromagnetic radiation spectrum. Light waves differ based on their frequency (wavelength = hue) and amplitude (intensity = brightness).

In the vertebrate retina, there are two types of light receptors (photoreceptors): cones and rods. Cones, which are the source of color vision, exist in three forms—L, M, and S—and they are differentially sensitive to different wavelengths. Cones are located in the retina, along with the dim-light, achromatic receptors (rods). Cones are found in the fovea, the central region of the retina, whereas rods are found in the peripheral regions of the retina.

Visual signals travel from the eye over the axons of retinal ganglion cells, which make up the optic nerves. Ganglion cells come in several versions. Some ganglion cell axons carry information on form, movement, depth, and brightness, while other axons carry information on color and fine detail. Visual information is sent to the superior colliculi in the midbrain, where coordination of eye movements and integration of auditory information takes place. Visual information is also sent to the suprachiasmatic nucleus (SCN) of the hypothalamus, which plays a role in the circadian cycle.

Exercises

1. Which of the following statements about mechanoreceptors is false?
    1. Pacini corpuscles are found in both glabrous and hairy skin.
    2. Merkel's disks are abundant on the fingertips and lips.
    3. Ruffini endings are encapsulated mechanoreceptors.
    4. Meissner's corpuscles extend into the lower dermis.
2. Why do people over 55 often need reading glasses?
    1. Their cornea no longer focuses correctly.
    2. Their lens no longer focuses correctly.
    3. Their eyeball has elongated with age, causing images to focus in front of their retina.
    4. Their retina has thinned with age, making vision more difficult.
3. Why is it easier to see images at night using peripheral, rather than the central, vision?
    1. Cones are denser in the periphery of the retina.
    2. Bipolar cells are denser in the periphery of the retina.
    3. Rods are denser in the periphery of the retina.
    4. The optic nerve exits at the periphery of the retina.
4. A person catching a ball must coordinate her head and eyes. What part of the brain is helping to do this?
5. How could the pineal gland, the brain structure that plays a role in annual cycles, use visual information from the suprachiasmatic nucleus of the hypothalamus?
6. How is the relationship between photoreceptors and bipolar cells different from other sensory receptors and adjacent cells?

Answers

1. D
2. B
3. C
4. D
5. The pineal gland could use length-of-day information to determine the time of year, for example. Day length is shorter in the winter than it is in the summer. For many animals and plants, photoperiod cues them to reproduce at a certain time of year.
6. The photoreceptors tonically inhibit the bipolar cells, and stimulation of the receptors turns this inhibition off, activating the bipolar cells.

Glossary

circadian
describes a time cycle about one day in length

cone
weakly photosensitive, chromatic, cone-shaped neuron in the fovea of the retina that detects bright light and is used in daytime color vision

cornea
transparent layer over the front of the eye that helps focus light waves

fovea
region in the center of the retina with a high density of photoreceptors and which is responsible for acute vision

hyperopia
(also, farsightedness) visual defect in which the image focus falls behind the retina, thereby making images in the distance clear, but close-up images blurry

iris
pigmented, circular muscle at the front of the eye that regulates the amount of light entering the eye

lens
transparent, convex structure behind the cornea that helps focus light waves on the retina

myopia
(also, nearsightedness) visual defect in which the image focus falls in front of the retina, thereby making images in the distance blurry, but close-up images clear
in the basilar membrane, the site of the transduction of sound, a mechanical wave, to a neural signal

presbyopia
visual defect in which the image focus falls behind the retina, thereby making images in the distance clear, but close-up images blurry; caused by age-based changes in the lens

rhodopsin
main photopigment in vertebrates

rod
strongly photosensitive, achromatic, cylindrical neuron in the outer edges of the retina that detects dim light and is used in

superior colliculus
paired structure in the top of the midbrain, which manages eye movements and auditory integration

suprachiasmatic nucleus
cluster of cells in the hypothalamus that plays a role in the circadian cycle

tonic activity
in a neuron, slight continuous activity while at rest

vision
sense of sight

Chapter 18. The Endocrine System

Figure 18.1.
The process of amphibian metamorphosis, as seen in the tadpole-to-frog stages shown here, is driven by hormones. (credit "tadpole": modification of work by Brian Gratwicke)

Introduction

An animal's endocrine system controls body processes through the production, secretion, and regulation of hormones, which serve as chemical "messengers" functioning in cellular and organ activity and, ultimately, maintaining the body's homeostasis. The endocrine system plays a role in growth, metabolism, and sexual development. In humans, common endocrine system diseases include thyroid disease and diabetes mellitus. In organisms that undergo metamorphosis, the process is controlled by the endocrine system. The transformation from tadpole to frog, for example, is complex and nuanced to adapt to specific environments and ecological circumstances.

18.1 Types of Hormones

Learning Objectives

By the end of this section, you will be able to:

- List the different types of hormones
- Explain their role in maintaining homeostasis

Maintaining homeostasis within the body requires the coordination of many different systems and organs. Communication between neighboring cells, and between cells and tissues in distant parts of the body, occurs through the release of chemicals called hormones. Hormones are released into body fluids (usually blood) that carry these chemicals to their target cells. At the target cells, which are cells that have a receptor for a signal or ligand from a signal cell, the hormones elicit a response. The cells, tissues, and organs that secrete hormones make up the endocrine system. Examples of glands of the endocrine system include the adrenal glands, which produce hormones such as epinephrine and norepinephrine that regulate responses to stress, and the thyroid gland, which produces thyroid hormones that regulate metabolic rates.

Although there are many different hormones in the human body, they can be divided into three classes based on their chemical structure: lipid-derived, amino acid-derived, and peptide (peptide and proteins) hormones. One of the key distinguishing features of lipid-derived hormones is that they can diffuse across plasma membranes whereas the amino acid-derived and 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 18.2. The primary class of lipid hormones in humans is the steroid hormones. Chemically, these hormones are usually ketones or alcohols; their chemical names will end in “-ol” for alcohols or “-one” for ketones. 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 they are transported by transport proteins in blood. As a result, they remain in circulation longer than peptide hormones. For example, cortisol has a half-life of 60 to 90 minutes, while epinephrine, an amino acid derived-hormone, has a half-life of approximately one minute.

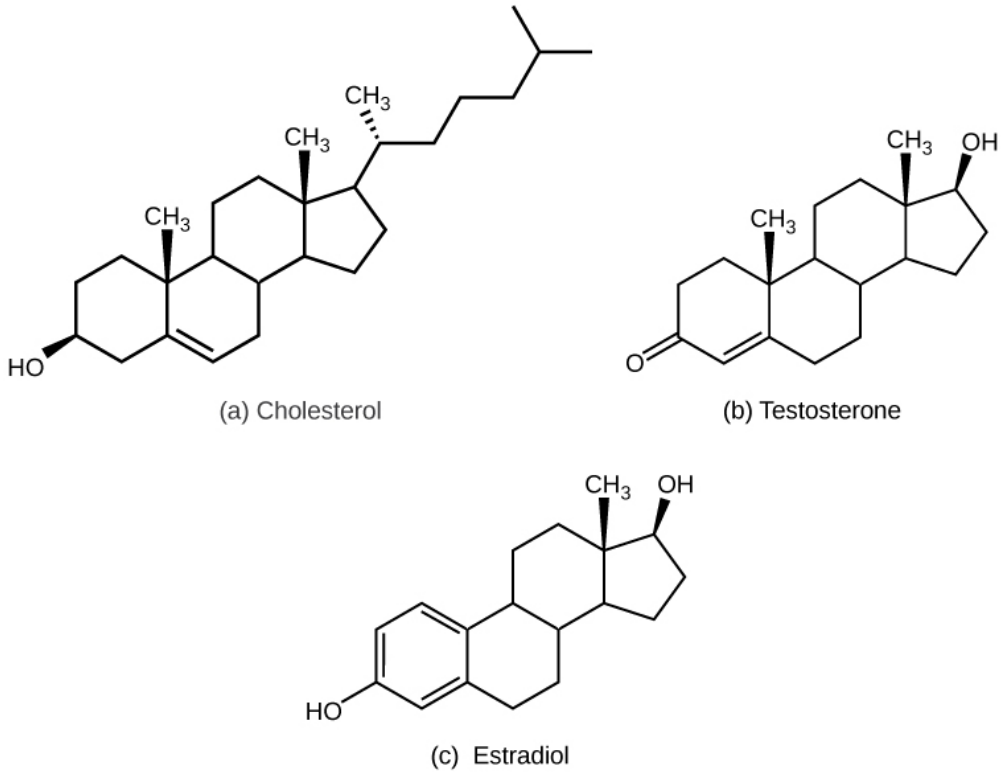


Figure 18.2. The structures shown here represent (a) cholesterol, plus the steroid hormones (b) testosterone and (c) estradiol.

Amino Acid-Derived Hormones

The **amino acid-derived hormones** are relatively small molecules that are derived from the amino acids tyrosine and tryptophan, shown in Figure 18.3. If a hormone is amino acid-derived, its chemical name will end in “-ine”. 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.

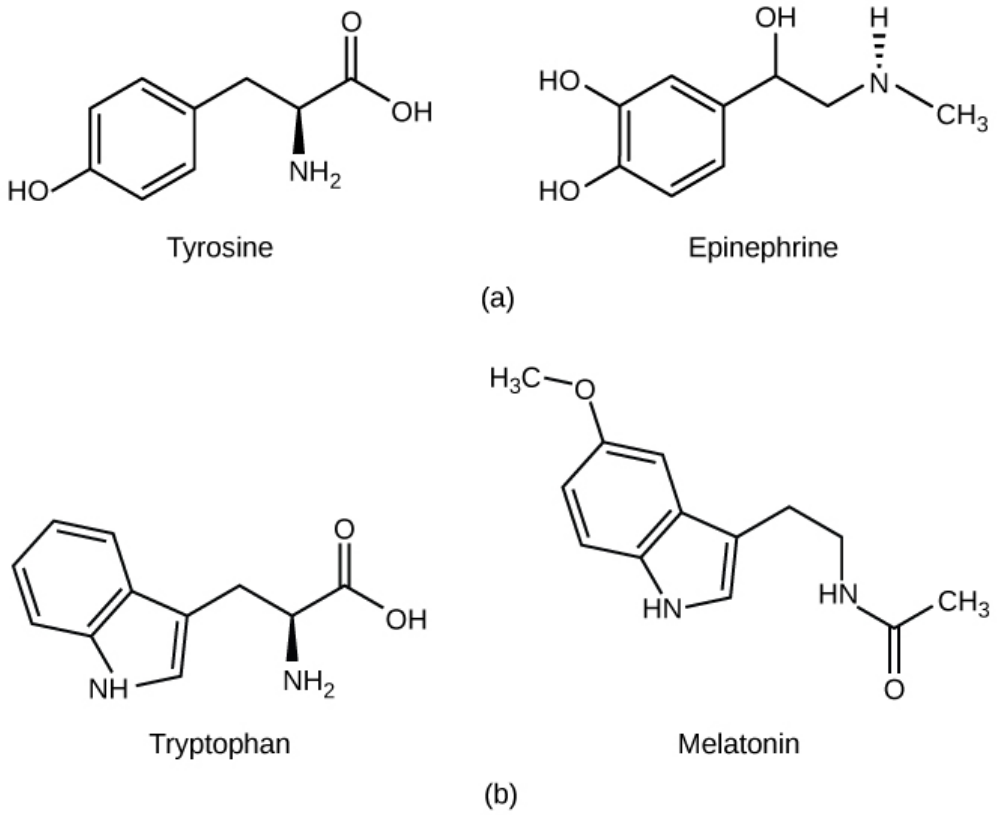


Figure 18.3. (a) The hormone epinephrine, which triggers the fight-or-flight response, is derived from the amino acid tyrosine. (b) The hormone melatonin, which regulates circadian rhythms, is derived from the amino acid tryptophan.

Peptide Hormones

The structure of **peptide hormones** is that of a polypeptide chain (chain of amino acids). The peptide hormones include molecules that are short polypeptide chains, such as antidiuretic hormone and oxytocin produced in the brain and released into the blood in the posterior pituitary gland. This class also includes small proteins, like growth hormones produced by the pituitary, and large glycoproteins such as follicle-stimulating hormone produced by the pituitary. Figure 18.4 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 and insoluble in lipids. These hormones cannot pass through plasma membranes of cells; therefore, their receptors are found on the surface of the target cells.

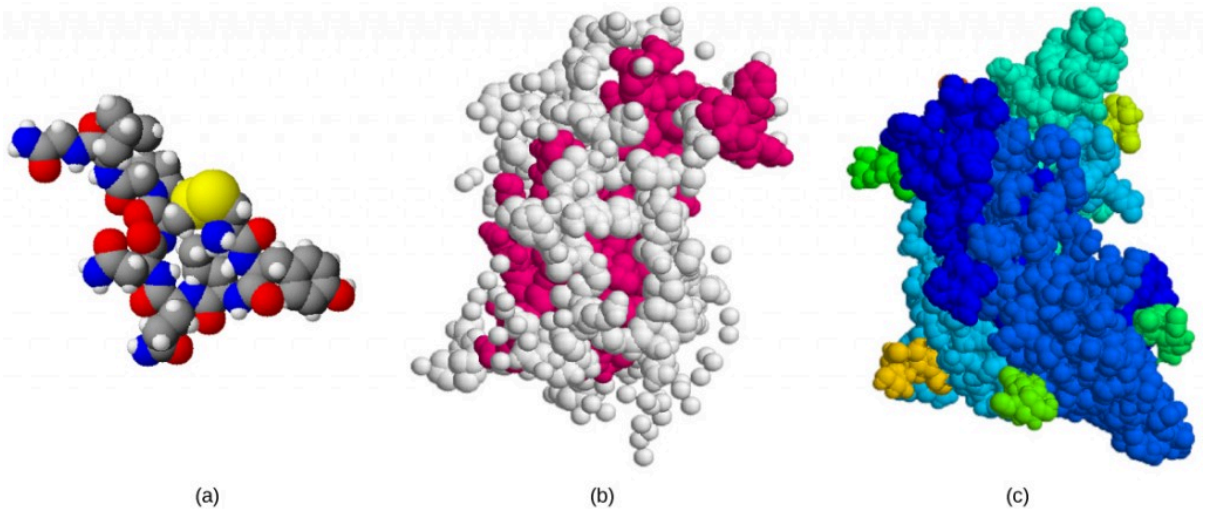


Figure 18.4. The structures of peptide hormones (a) oxytocin, (b) growth hormone, and (c) follicle-stimulating hormone are shown. These peptide hormones are much larger than those derived from cholesterol or amino acids.

Endocrinologist

An endocrinologist is a medical doctor who specializes in treating disorders of the endocrine glands, hormone systems, and glucose and lipid metabolic pathways. An endocrine surgeon specializes in the surgical treatment of endocrine diseases and glands. Some of the diseases that are managed by endocrinologists: disorders of the pancreas (diabetes mellitus), disorders of the pituitary (gigantism, acromegaly, and pituitary dwarfism), disorders of the thyroid gland (goiter and Graves' disease), and disorders of the adrenal glands (Cushing's disease and Addison's disease).

Endocrinologists are required to assess patients and diagnose endocrine disorders through extensive use of laboratory tests. Many endocrine diseases are diagnosed using tests that stimulate or suppress endocrine organ functioning. Blood samples are then drawn to determine the effect of stimulating or suppressing an endocrine organ on the production of hormones. For example, to diagnose diabetes mellitus, patients are required to fast for 12 to 24 hours. They are then given a sugary drink, which stimulates the pancreas to produce insulin to decrease blood glucose levels. A blood sample is taken one to two hours after the sugar drink is consumed. If the pancreas is functioning properly, the blood glucose level will be within a normal range. Another example is the A1C test, which can be performed during blood screening. The A1C test measures average blood glucose levels over the past two to three months by examining how well the blood glucose is being managed over a long time.

Once a disease has been diagnosed, endocrinologists can prescribe lifestyle changes and/or medications to treat the disease. Some cases of diabetes mellitus can be managed by exercise, weight loss, and a healthy diet; in other cases, medications may be required to enhance insulin release. If the disease cannot be controlled by these means, the endocrinologist may prescribe insulin injections.

In addition to clinical practice, endocrinologists may also be involved in primary research and development activities. For example, ongoing islet transplant research is investigating how healthy pancreas islet cells may be transplanted into diabetic patients. Successful islet transplants may allow patients to stop taking insulin injections.

Summary

There are three basic types of hormones: lipid-derived, amino acid-derived, and peptide. Lipid-derived hormones are structurally similar to cholesterol and include steroid hormones such as estradiol and testosterone. Amino acid-derived hormones are relatively small molecules and include the adrenal hormones epinephrine and norepinephrine. Peptide hormones are polypeptide chains or proteins and include the pituitary hormones, antidiuretic hormone (vasopressin), and oxytocin.

Exercises

1. A newly discovered hormone contains four amino acids linked together. Under which chemical class would this hormone be classified?
    1. libid-derived hormone
    2. amino acid-derived hormone
    3. peptide hormone
    4. glycoprotien
2. Which class of hormones can diffuse through plasma membranes?
    1. libid-derived hormones
    2. amino acid-derived hormones
    3. peptide hormones
    4. glycoprotien hormones
3. Although there are many different hormones in the human body, they can be divided into three classes based on their chemical structure. What are these classes and what is one factor that distinguishes them?
4. Where is insulin stored, and why would it be released?

Answers

1. C
2. A
3. Although there are many different hormones in the human body, they can be divided into three classes based on their chemical structure: lipid-derived, amino acid-derived, and peptide hormones. One of the key distinguishing features of the lipid-derived hormones is that they can diffuse across plasma membranes whereas the amino acid-derived and peptide hormones cannot.
4. Secreted peptides such as 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.

Glossary

amino acid-derived hormone
hormone derived from amino acids

estrogens
– a group of steroid hormones, including estradiol and several others, that are produced by the ovaries and elicit

lipid-derived hormone
hormone derived mostly from cholesterol

peptide hormone
hormone composed of a polypeptide chain

18.2 How Hormones Work

Learning Objectives

By the end of this section, you will be able to:

- Explain how hormones work
- Discuss the role of different types of hormone receptors

Hormones mediate changes in target cells by binding to specific **hormone receptors**. In this way, even though hormones circulate throughout the body and come into contact with many different cell types, they only affect cells that possess the necessary receptors. Receptors for a specific hormone may be found on many different cells or may be limited to a small number of specialized cells. For example, thyroid hormones act on many different tissue types, stimulating metabolic activity throughout the body. Cells can have many receptors for the same hormone but often also possess receptors for different types of hormones. The number of receptors that respond to a hormone determines the cell's sensitivity to that hormone, and the resulting cellular response. Additionally, the number of receptors that respond to a hormone can change over time, resulting in increased or decreased cell sensitivity. In **up-regulation**, the number of receptors increases in response to rising hormone levels, making the cell more sensitive to the hormone and allowing for more cellular activity. When the number of receptors decreases in response to rising hormone levels, called **down-regulation**, cellular activity is reduced.

Receptor binding alters cellular activity and results in an increase or decrease in normal body processes. Depending on the location of the protein receptor on the target cell and the chemical structure of the hormone, hormones can mediate changes directly by binding to **intracellular hormone receptors** and modulating gene transcription, or indirectly by binding to cell surface receptors and stimulating signaling pathways.

Intracellular Hormone Receptors

Lipid-derived (soluble) hormones such as steroid hormones diffuse across the membranes of the endocrine cell. Once outside the cell, they bind to transport proteins that keep them soluble in the bloodstream. At the target cell, the hormones are released from the carrier protein and diffuse across the lipid bilayer of the plasma membrane of cells. The steroid hormones pass through the plasma membrane of a target cell and adhere to intracellular receptors residing in the cytoplasm or in the nucleus. The cell signaling pathways induced by the steroid hormones regulate specific genes on the cell's DNA. The hormones and receptor complex act as transcription regulators by increasing or decreasing the synthesis of mRNA molecules of specific genes. This, in turn, determines the amount of corresponding protein that is synthesized by altering gene expression. This protein can be used either to change the structure of the cell or to produce enzymes that catalyze chemical reactions. In this way, the steroid hormone regulates specific cell processes as illustrated in Figure 18.5.

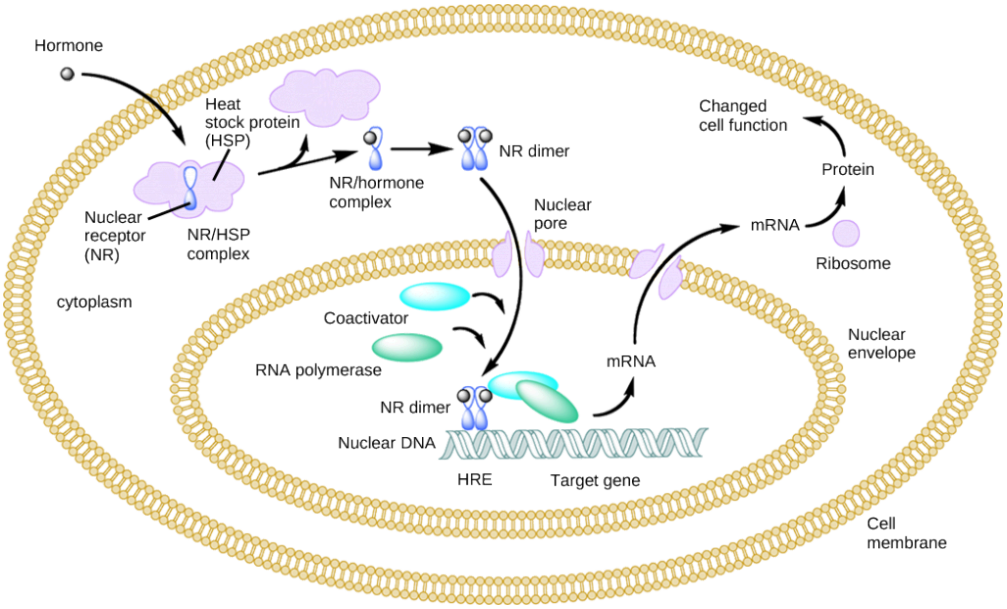


Figure 18.5. An intracellular nuclear receptor (NR) is located in the cytoplasm bound to a heat shock protein (HSP). Upon hormone binding, the receptor dissociates from the heat shock protein and translocates to the nucleus. In the nucleus, the hormone-receptor complex binds to a DNA sequence called a hormone response element (HRE), which triggers gene transcription and translation. The corresponding protein product can then mediate changes in cell function.

Heat shock proteins (HSP) are so named because they help refold misfolded proteins. In response to increased temperature (a “heat shock”), heat shock proteins are activated by release from the NR/HSP complex. At the same time, transcription of HSP genes is activated. Why do you think the cell responds to a heat shock by increasing the activity of proteins that help refold misfolded proteins?

Other lipid-soluble hormones that are not steroid hormones, such as vitamin D and thyroxine, have receptors located in the nucleus. The hormones diffuse across both the plasma membrane and the nuclear envelope, then bind to receptors in the nucleus. The hormone-receptor complex stimulates transcription of specific genes.

Plasma Membrane Hormone Receptors

Amino acid derived hormones and polypeptide hormones are not lipid-derived (lipid-soluble) and therefore cannot diffuse through the plasma membrane of cells. Lipid insoluble hormones bind to receptors on the outer surface of the plasma membrane, via **plasma membrane hormone receptors**. Unlike steroid hormones, lipid insoluble hormones do not directly affect the target cell because they cannot enter the cell and act directly on DNA. Binding of these hormones to a cell surface receptor results in activation of a signaling pathway; this triggers intracellular activity and carries out the specific effects associated with the hormone. In this way, nothing passes through the cell membrane; the hormone that binds at the surface remains at the surface of the cell while the intracellular product remains inside the cell. The hormone that initiates the signaling pathway is called a **first messenger**, which activates a second messenger in the cytoplasm, as illustrated in Figure 18.6.

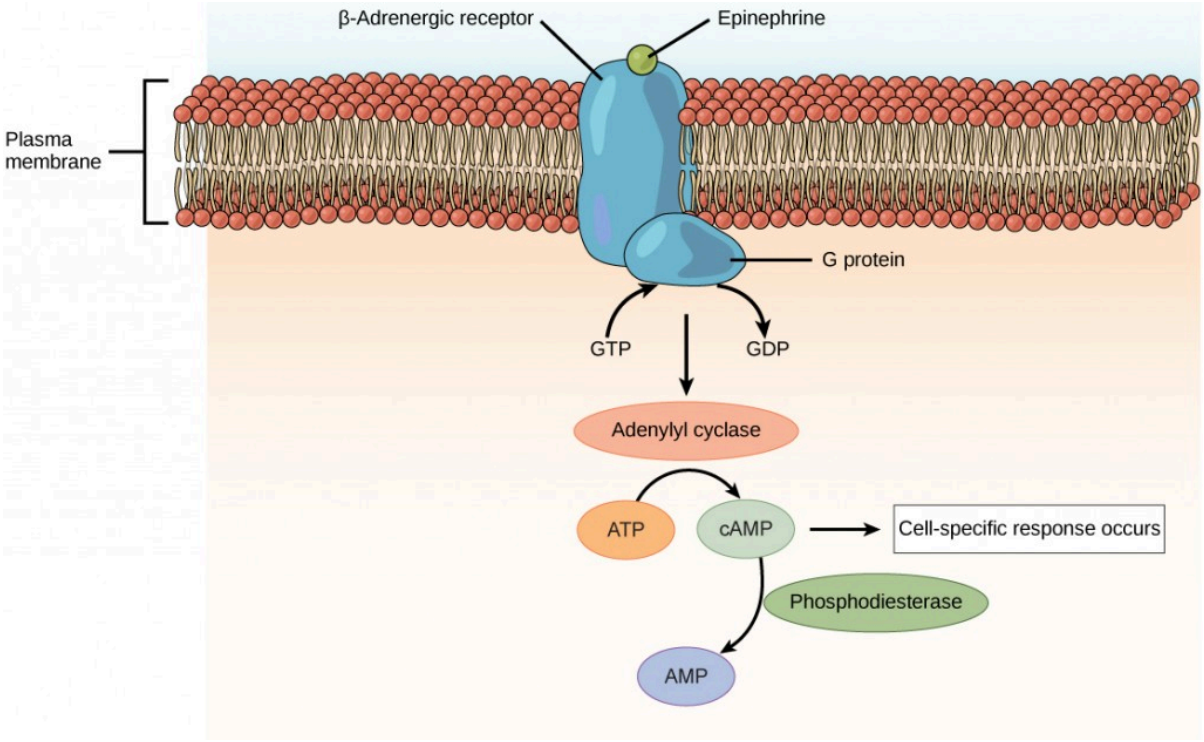


Figure 18.6. The amino acid-derived hormones epinephrine and norepinephrine bind to beta-adrenergic receptors on the plasma membrane of cells. Hormone binding to receptor activates a G-protein, which in turn activates adenylyl cyclase, converting ATP to cAMP. cAMP is a second messenger that mediates a cell-specific response. An enzyme called phosphodiesterase breaks down cAMP, terminating the signal.

One very important second messenger is cyclic AMP (cAMP). When a hormone binds to its membrane receptor, a **G-protein** that is associated with the receptor is activated; G-proteins are proteins separate from receptors that are found in the cell membrane. When a hormone is not bound to the receptor, the G-protein is inactive and is bound to guanosine diphosphate, or GDP. When a hormone binds to the receptor, the G-protein is activated by binding guanosine triphosphate, or GTP, in place of GDP. After binding, GTP is hydrolysed by the G-protein into GDP and becomes inactive.

The activated G-protein in turn activates a membrane-bound enzyme called **adenylyl cyclase**. Adenylyl cyclase catalyzes the conversion of ATP to cAMP. cAMP, in turn, activates a group of proteins called protein kinases, which transfer a phosphate group from ATP to a substrate molecule in a process called phosphorylation. The phosphorylation of a substrate molecule changes its structural orientation, thereby activating it. These activated molecules can then mediate changes in cellular processes.

The effect of a hormone is amplified as the signaling pathway progresses. The binding of a hormone at a single receptor causes the activation of many G-proteins, which activates adenylyl cyclase. Each molecule of adenylyl cyclase then triggers the formation of many molecules of cAMP. Further amplification occurs as protein kinases, once activated by cAMP, can catalyze many reactions. In this way, a small amount of hormone can trigger the formation of a large amount of cellular product. To stop hormone activity, cAMP is deactivated by the cytoplasmic enzyme **phosphodiesterase**, or PDE. PDE is always present in the cell and breaks down cAMP to control hormone activity, preventing overproduction of cellular products.

The specific response of a cell to a lipid insoluble hormone depends on the type of receptors that are

present on the cell membrane and the substrate molecules present in the cell cytoplasm. Cellular responses to hormone binding of a receptor include altering membrane permeability and metabolic pathways, stimulating synthesis of proteins and enzymes, and activating hormone release.

Summary

Hormones cause cellular changes by binding to receptors on target cells. The number of receptors on a target cell can increase or decrease in response to hormone activity. Hormones can affect cells directly through intracellular hormone receptors or indirectly through plasma membrane hormone receptors.

Lipid-derived (soluble) hormones can enter the cell by diffusing across the plasma membrane and binding to DNA to regulate gene transcription and to change the cell's activities by inducing production of proteins that affect, in general, the long-term structure and function of the cell. Lipid insoluble hormones bind to receptors on the plasma membrane surface and trigger a signaling pathway to change the cell's activities by inducing production of various cell products that affect the cell in the short-term. The hormone is called a first messenger and the cellular component is called a second messenger. G-proteins activate the second messenger (cyclic AMP), triggering the cellular response. Response to hormone binding is amplified as the signaling pathway progresses. Cellular responses to hormones include the production of proteins and enzymes and altered membrane permeability.

Exercises

1. A new antagonist molecule has been discovered that binds to and blocks plasma membrane receptors. What effect will this antagonist have on testosterone, a steroid hormone?
    1. It will block testosterone from binding to its receptor.
    2. It will block testosterone from activating cAMP signaling.
    3. It will increase testosterone-mediated signaling.
    4. It will not affect testosterone-mediated signaling.
2. What effect will a cAMP inhibitor have on a peptide hormone-mediated signaling pathway?
    1. It will prevent the hormone from binding its receptor.
    2. It will prevent activation of a G-protein.
    3. It will prevent activation of adenylate cyclase.
    4. It will prevent activation of protein kinases.
3. Name two important functions of hormone receptors.
4. How can hormones mediate changes?

Answers

1. D
2. D
3. The number of receptors that respond to a hormone can change, resulting in increased or

decreased cell sensitivity. The number of receptors can increase in response to rising hormone levels, called up-regulation, making the cell more sensitive to the hormone and allowing for more cellular activity. The number of receptors can also decrease in response to rising hormone levels, called down-regulation, leading to reduced cellular activity.

4. 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 receptors and modulating gene transcription, or indirectly by binding to cell surface receptors and stimulating signaling pathways.

Glossary

adenylate cyclase
an enzyme that catalyzes the conversion of ATP to cyclic AMP

down-regulation
a decrease in the number of hormone receptors in response to increased hormone levels

first messenger
the hormone that binds to a plasma membrane hormone receptor to trigger a signal transduction pathway

G-protein
a membrane protein activated by the hormone first messenger to activate formation of cyclic AMP

hormone receptor
the cellular protein that binds to a hormone

intracellular hormone receptor
a hormone receptor in the cytoplasm or nucleus of a cell

phosphodiesterase (PDE)
enzyme that deactivates cAMP, stopping hormone activity

plasma membrane hormone receptor
a hormone receptor on the surface of the plasma membrane of a cell

up-regulation
an increase in the number of hormone receptors in response to increased hormone levels

18.3 Regulation of Body Processes

Learning Objectives

By the end of this section, you will be able to:

- Explain how hormones regulate the excretory system
- Discuss the role of hormones in the reproductive system
- Describe how hormones regulate metabolism
- Explain the role of hormones in different diseases

Hormones have a wide range of effects and modulate many different body processes. The key regulatory processes that will be examined here are those affecting the excretory system, the reproductive system, metabolism, blood calcium concentrations, growth, and the stress response.

Hormonal Regulation of the Excretory System

Maintaining a proper water balance in the body is important to avoid dehydration or over-hydration (hyponatremia). The water concentration of the body is monitored by **osmoreceptors** in the hypothalamus, which detect the concentration of electrolytes in the extracellular fluid. The concentration of electrolytes in the blood rises when there is water loss caused by excessive perspiration, inadequate water intake, or low blood volume due to blood loss. An increase in blood electrolyte levels results in a neuronal signal being sent from the osmoreceptors in hypothalamic nuclei. The pituitary gland has two components: anterior and posterior. The anterior pituitary is composed of glandular cells that secrete protein hormones. The posterior pituitary is an extension of the hypothalamus. It is composed largely of neurons that are continuous with the hypothalamus.

The hypothalamus produces a polypeptide hormone known as **antidiuretic hormone (ADH)**, which is transported to and released from the posterior pituitary gland. The principal action of ADH is to regulate the amount of water excreted by the kidneys. As ADH (which is also known as vasopressin) causes direct water reabsorption from the kidney tubules, salts and wastes are concentrated in what will eventually be excreted as urine. The hypothalamus controls the mechanisms of ADH secretion, either by regulating blood volume or the concentration of water in the blood. Dehydration or physiological stress can cause an increase of osmolarity above 300 mOsm/L, which in turn, raises ADH secretion and water will be retained, causing an increase in blood pressure. ADH travels in the bloodstream to the kidneys. Once at the kidneys, ADH changes the kidneys to become more permeable to water by temporarily inserting water channels, aquaporins, into the kidney tubules. Water moves out of the kidney tubules through the aquaporins, reducing urine volume. The water is reabsorbed into the capillaries lowering blood osmolarity back toward normal. As blood osmolarity decreases, a negative feedback mechanism reduces osmoreceptor activity in the hypothalamus, and ADH secretion is

reduced. ADH release can be reduced by certain substances, including alcohol, which can cause increased urine production and dehydration.

Chronic underproduction of ADH or a mutation in the ADH receptor results in **diabetes insipidus**. If the posterior pituitary does not release enough ADH, water cannot be retained by the kidneys and is lost as urine. This causes increased thirst, but water taken in is lost again and must be continually consumed. If the condition is not severe, dehydration may not occur, but severe cases can lead to electrolyte imbalances due to dehydration.

Another hormone responsible for maintaining electrolyte concentrations in extracellular fluids is **aldosterone**, a steroid hormone that is produced by the adrenal cortex. In contrast to ADH, which promotes the reabsorption of water to maintain proper water balance, aldosterone maintains proper water balance by enhancing Na^+ reabsorption and K^+ secretion from extracellular fluid of the cells in kidney tubules. Because it is produced in the cortex of the adrenal gland and affects the concentrations of minerals Na^+ and K^+, aldosterone is referred to as a **mineralocorticoid**, a corticosteroid that affects ion and water balance. Aldosterone release is stimulated by a decrease in blood sodium levels, blood volume, or blood pressure, or an increase in blood potassium levels. It also prevents the loss of Na^+ from sweat, saliva, and gastric juice. The reabsorption of Na^+ also results in the osmotic reabsorption of water, which alters blood volume and blood pressure.

Aldosterone production can be stimulated by low blood pressure, which triggers a sequence of chemical release, as illustrated in Figure 18.7. When blood pressure drops, the renin-angiotensin-aldosterone system (RAAS) is activated. Cells in the juxtaglomerular apparatus, which regulates the functions of the nephrons of the kidney, detect this and release **renin**. Renin, an enzyme, circulates in the blood and reacts with a plasma protein produced by the liver called angiotensinogen. When angiotensinogen is cleaved by renin, it produces angiotensin I, which is then converted into angiotensin II in the lungs. Angiotensin II functions as a hormone and then causes the release of the hormone aldosterone by the adrenal cortex, resulting in increased Na^+ reabsorption, water retention, and an increase in blood pressure. Angiotensin II in addition to being a potent vasoconstrictor also causes an increase in ADH and increased thirst, both of which help to raise blood pressure.

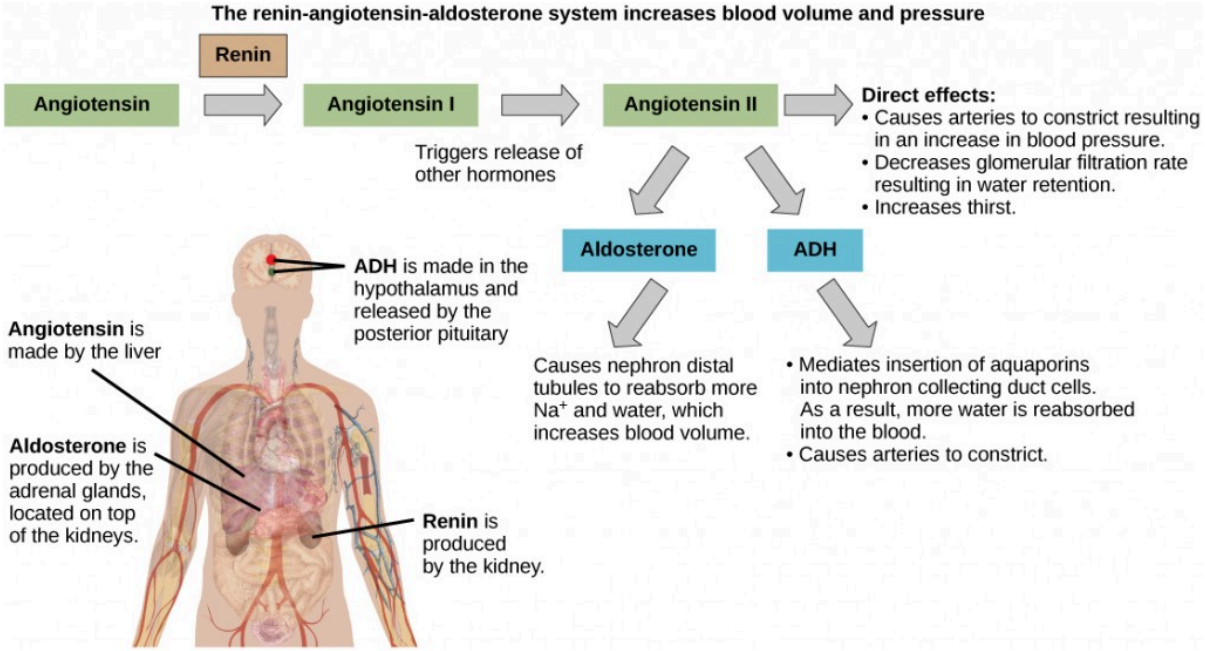


Figure 18.7.
ADH and aldosterone increase blood pressure and volume. Angiotensin II stimulates release of these hormones. Angiotensin II, in turn, is formed when renin cleaves angiotensin. (credit: modification of work by Mikael Häggström)

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. The adrenal cortex also produces small amounts of testosterone precursor, although the role of this additional hormone production is not fully understood.

The Dangers of Synthetic Hormones

Figure 18.8.
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 18.9. 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.

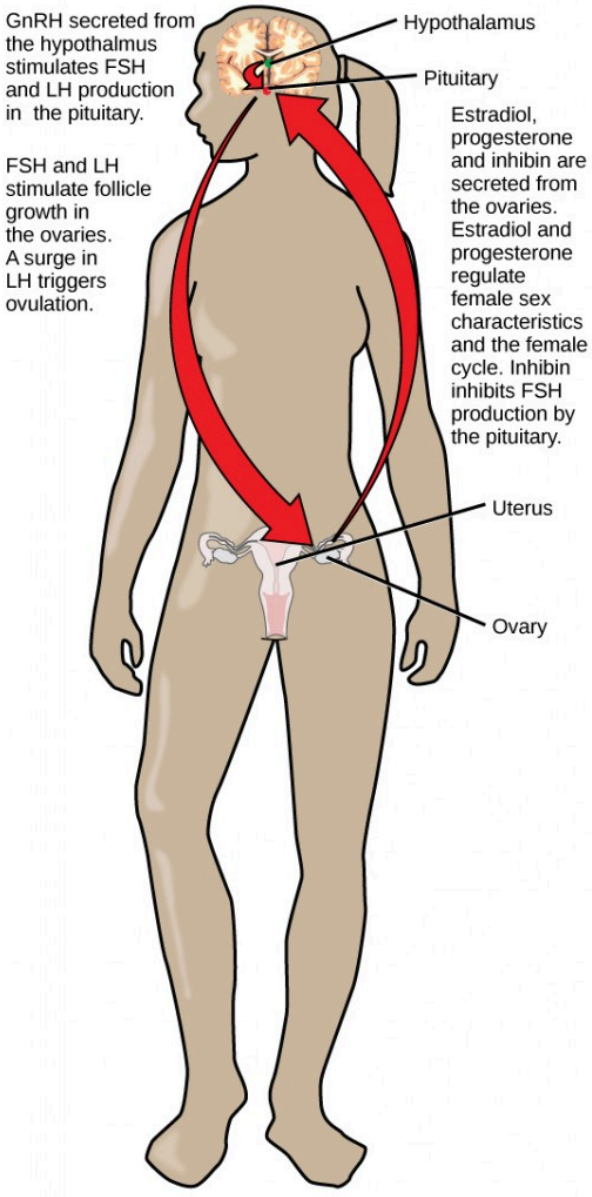


Figure 18.9. Hormonal regulation of the female reproductive system involves hormones from the hypothalamus, pituitary, and ovaries.

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.

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 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 and the liver. This results from an insulin-mediated increase in the number of glucose transporter proteins in cell 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. This allows glucose to enter the cell, where it can be used as an energy source. However, this does not occur in all cells: some cells, including those in the kidneys and brain, can access glucose without the use of insulin. 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.

Concept in Action

This animation describe the role of insulin and the pancreas in diabetes.

Impaired insulin function can lead to a condition called **diabetes mellitus**, the main symptoms of which are illustrated in Figure 18.10. 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.

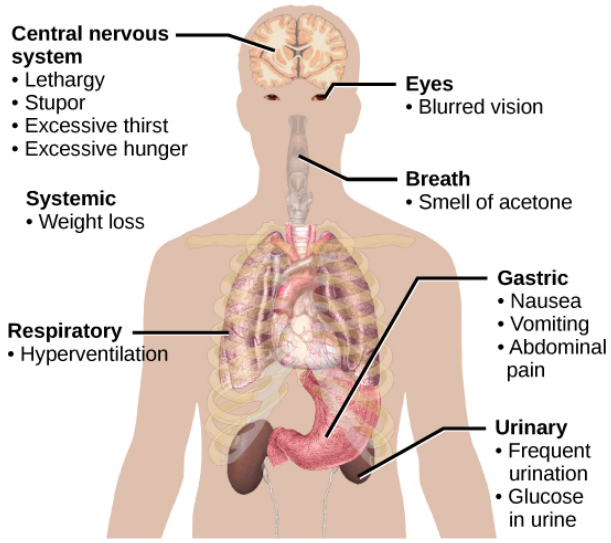


Figure 18.10.
The main symptoms of diabetes are shown. (credit: modification of work by Mikael Häggström)

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 Figure 18.11.

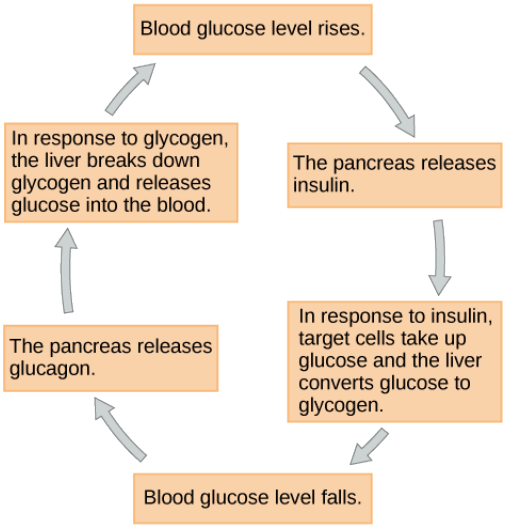


Figure 18.11.
Insulin and glucagon regulate blood glucose levels.

Pancreatic tumors may cause excess secretion of glucagon. Type I diabetes results from the failure of the pancreas to produce insulin. Which of the following statement about these two conditions is true?

1. A pancreatic tumor and type I diabetes will have the opposite effects on blood sugar levels.
2. A pancreatic tumor and type I diabetes will both cause hyperglycemia.
3. A pancreatic tumor and type I diabetes will both cause hypoglycemia.
4. Both pancreatic tumors and type I diabetes result in the inability of cells to take up glucose.

Regulation of Blood Glucose Levels by Thyroid Hormones

The basal metabolic rate, which is the amount of calories required by the body at rest, is determined by two hormones produced by the thyroid gland: **thyroxine**, also known as tetraiodothyronine or T_4, and **triiodothyronine**, also known as T_3. These hormones affect nearly every cell in the body except for the adult brain, uterus, testes, blood cells, and spleen. They are transported across the plasma membrane of target cells and bind to receptors on the mitochondria resulting in increased ATP production. In the nucleus, T_3 and T_4 activate genes involved in energy production and glucose oxidation. This results in increased rates of metabolism and body heat production, which is known as the hormone's calorigenic effect.

T_3 and T_4 release from the thyroid gland is stimulated by **thyroid-stimulating hormone (TSH)**, which is produced by the anterior pituitary. TSH binding at the receptors of the follicle of the thyroid triggers the production of T_3 and T_4 from a glycoprotein called **thyroglobulin**. Thyroglobulin is present in the follicles of the thyroid, and is converted into thyroid hormones with the addition of iodine. Iodine is formed from iodide ions that are actively transported into the thyroid follicle from the bloodstream. A peroxidase enzyme then attaches the iodine to the tyrosine amino acid found in thyroglobulin. T_3 has three iodine ions attached, while T_4 has four iodine ions attached. T_3 and T_4 are then released into the bloodstream, with T_4 being released in much greater amounts than T_3. As T_3 is more active than T_4 and

is responsible for most of the effects of thyroid hormones, tissues of the body convert T_4 to T_3 by the removal of an iodine ion. Most of the released T_3 and T_4 becomes attached to transport proteins in the bloodstream and is unable to cross the plasma membrane of cells. These protein-bound molecules are only released when blood levels of the unattached hormone begin to decline. In this way, a week's worth of reserve hormone is maintained in the blood. Increased T_3 and T_4 levels in the blood inhibit the release of TSH, which results in lower T_3 and T_4 release from the thyroid.

The follicular cells of the thyroid require iodides (anions of iodine) in order to synthesize T_3 and T_4. Iodides obtained from the diet are actively transported into follicle cells resulting in a concentration that is approximately 30 times higher than in blood. The typical diet in North America provides more iodine than required due to the addition of iodide to table salt. Inadequate iodine intake, which occurs in many developing countries, results in an inability to synthesize T_3 and T_4 hormones. The thyroid gland enlarges in a condition called **goiter**, which is caused by overproduction of TSH without the formation of thyroid hormone. Thyroglobulin is contained in a fluid called colloid, and TSH stimulation results in higher levels of colloid accumulation in the thyroid. In the absence of iodine, this is not converted to thyroid hormone, and colloid begins to accumulate more and more in the thyroid gland, leading to goiter.

Disorders can arise from both the underproduction and overproduction of thyroid hormones. **Hypothyroidism**, underproduction of the thyroid hormones, can cause a low metabolic rate leading to weight gain, sensitivity to cold, and reduced mental activity, among other symptoms. In children, hypothyroidism can cause cretinism, which can lead to mental retardation and growth defects. **Hyperthyroidism**, the overproduction of thyroid hormones, can lead to an increased metabolic rate and its effects: weight loss, excess heat production, sweating, and an increased heart rate. Graves' disease is one example of a hyperthyroid condition.

Hormonal Control of Blood Calcium Levels

Regulation of blood calcium concentrations is important for generation of muscle contractions and nerve impulses, which are electrically stimulated. If calcium levels get too high, membrane permeability to sodium decreases and membranes become less responsive. If calcium levels get too low, membrane permeability to sodium increases and convulsions or muscle spasms can result.

Blood calcium levels are regulated by **parathyroid hormone (PTH)**, which is produced by the parathyroid glands, as illustrated in Figure 18.12. PTH is released in response to low blood Ca^{2+} levels. PTH increases Ca^{2+} levels by targeting the skeleton, the kidneys, and the intestine. In the skeleton, PTH stimulates osteoclasts, which causes bone to be reabsorbed, releasing Ca^{2+} from bone into the blood. PTH also inhibits osteoblasts, reducing Ca^{2+} deposition in bone. In the intestines, PTH increases dietary Ca^{2+} absorption, and in the kidneys, PTH stimulates reabsorption of the CA^{2+}. While PTH acts directly on the kidneys to increase Ca^{2+} reabsorption, its effects on the intestine are indirect. PTH triggers the formation of calcitriol, an active form of vitamin D, which acts on the intestines to increase absorption of dietary calcium. PTH release is inhibited by rising blood calcium levels.

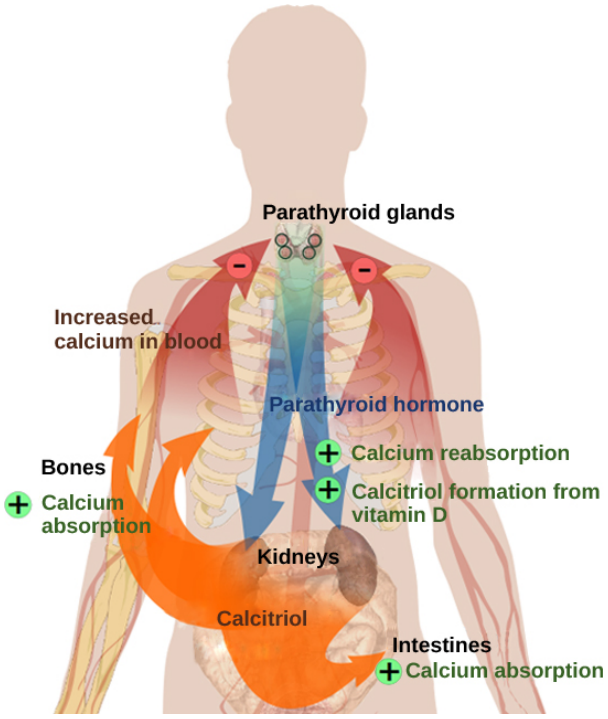


Figure 18.12.
Parathyroid hormone (PTH) is released in response to low blood calcium levels. It increases blood calcium levels by targeting the skeleton, the kidneys, and the intestine. (credit: modification of work by Mikael Häggström)

Hyperparathyroidism results from an overproduction of parathyroid hormone. This results in excessive calcium being removed from bones and introduced into blood circulation, producing structural weakness of the bones, which can lead to deformation and fractures, plus nervous system impairment due to high blood calcium levels. Hypoparathyroidism, the underproduction of PTH, results in extremely low levels of blood calcium, which causes impaired muscle function and may result in tetany (severe sustained muscle contraction).

The hormone **calcitonin**, which is produced by the parafollicular or C cells of the thyroid, has the opposite effect on blood calcium levels as does PTH. Calcitonin decreases blood calcium levels by inhibiting osteoclasts, stimulating osteoblasts, and stimulating calcium excretion by the kidneys. This results in calcium being added to the bones to promote structural integrity. Calcitonin is most important in children (when it stimulates bone growth), during pregnancy (when it reduces maternal bone loss), and during prolonged starvation (because it reduces bone mass loss). In healthy nonpregnant, unstarved adults, the role of calcitonin is unclear.

Hormonal Regulation of Growth

Hormonal regulation is required for the growth and replication of most cells in the body. **Growth hormone (GH)**, produced by the anterior portion of the pituitary gland, accelerates the rate of protein synthesis, particularly in skeletal muscle and bones. Growth hormone has direct and indirect mechanisms of action. The first direct action of GH is stimulation of triglyceride breakdown (lipolysis)

and release into the blood by adipocytes. This results in a switch by most tissues from utilizing glucose as an energy source to utilizing fatty acids. This process is called a **glucose-sparing effect**. In another direct mechanism, GH stimulates glycogen breakdown in the liver; the glycogen is then released into the blood as glucose. Blood glucose levels increase as most tissues are utilizing fatty acids instead of glucose for their energy needs. The GH mediated increase in blood glucose levels is called a **diabetogenic effect** because it is similar to the high blood glucose levels seen in diabetes mellitus.

The indirect mechanism of GH action is mediated by **insulin-like growth factors (IGFs)** or somatomedins, which are a family of growth-promoting proteins produced by the liver, which stimulates tissue growth. IGFs stimulate the uptake of amino acids from the blood, allowing the formation of new proteins, particularly in skeletal muscle cells, cartilage cells, and other target cells, as shown in Figure 18.13. This is especially important after a meal, when glucose and amino acid concentration levels are high in the blood. GH levels are regulated by two hormones produced by the hypothalamus. GH release is stimulated by **growth hormone-releasing hormone (GHRH)** and is inhibited by **growth hormone-inhibiting hormone (GHIH)**, also called somatostatin.

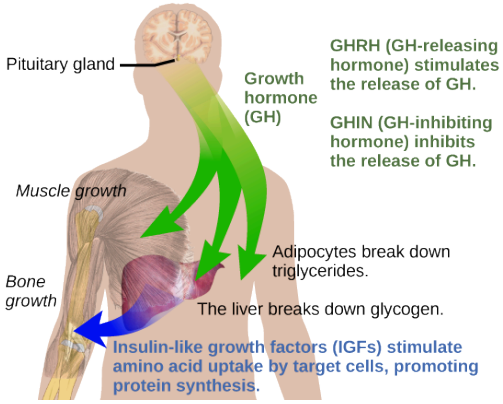


Figure 18.13.
Growth hormone directly accelerates the rate of protein synthesis in skeletal muscle and bones. Insulin-like growth factor 1 (IGF-1) is activated by growth hormone and also allows formation of new proteins in muscle cells and bone. (credit: modification of work by Mikael Häggström)

A balanced production of growth hormone is critical for proper development. Underproduction of GH in adults does not appear to cause any abnormalities, but in children it can result in **pituitary dwarfism**, in which growth is reduced. Pituitary dwarfism is characterized by symmetric body formation. In some cases, individuals are under 30 inches in height. Oversecretion of growth hormone can lead to **gigantism** in children, causing excessive growth. In some documented cases, individuals can reach heights of over eight feet. In adults, excessive GH can lead to **acromegaly**, a condition in which there is enlargement of bones in the face, hands, and feet that are still capable of growth.

Hormonal Regulation of Stress

When a threat or danger is perceived, the body responds by releasing hormones that will ready it for the

"fight-or-flight" response. The effects of this response are familiar to anyone who has been in a stressful situation: increased heart rate, dry mouth, and hair standing up.

Fight-or-Flight Response

Interactions of the endocrine hormones have evolved to ensure the body's internal environment remains stable. Stressors are stimuli that disrupt homeostasis. The sympathetic division of the vertebrate autonomic nervous system has evolved the fight-or-flight response to counter stress-induced disruptions of homeostasis. In the initial alarm phase, the sympathetic nervous system stimulates an increase in energy levels through increased blood glucose levels. This prepares the body for physical activity that may be required to respond to stress: to either fight for survival or to flee from danger.

However, some stresses, such as illness or injury, can last for a long time. Glycogen reserves, which provide energy in the short-term response to stress, are exhausted after several hours and cannot meet long-term energy needs. If glycogen reserves were the only energy source available, neural functioning could not be maintained once the reserves became depleted due to the nervous system's high requirement for glucose. In this situation, the body has evolved a response to counter long-term stress through the actions of the glucocorticoids, which ensure that long-term energy requirements can be met. The glucocorticoids mobilize lipid and protein reserves, stimulate gluconeogenesis, conserve glucose for use by neural tissue, and stimulate the conservation of salts and water. The mechanisms to maintain homeostasis that are described here are those observed in the human body. However, the fight-or-flight response exists in some form in all vertebrates.

The sympathetic nervous system regulates the stress response via the hypothalamus. Stressful stimuli cause the hypothalamus to signal the adrenal medulla (which mediates short-term stress responses) via nerve impulses, and the adrenal cortex, which mediates long-term stress responses, via the hormone **adrenocorticotropic hormone (ACTH)**, which is produced by the anterior pituitary.

Short-term Stress Response

When presented with a stressful situation, the body responds by calling for the release of hormones that provide a burst of energy. The hormones **epinephrine** (also known as adrenaline) and **norepinephrine** (also known as noradrenaline) are released by the adrenal medulla. How do these hormones provide a burst of energy? Epinephrine and norepinephrine increase blood glucose levels by stimulating the liver and skeletal muscles to break down glycogen and by stimulating glucose release by liver cells. Additionally, these hormones increase oxygen availability to cells by increasing the heart rate and dilating the bronchioles. The hormones also prioritize body function by increasing blood supply to essential organs such as the heart, brain, and skeletal muscles, while restricting blood flow to organs not in immediate need, such as the skin, digestive system, and kidneys. Epinephrine and norepinephrine are collectively called catecholamines.

Concept in Action

Watch this

Discovery Channel animation describing the flight-or-flight response.

Long-term Stress Response

Long-term stress response differs from short-term stress response. The body cannot sustain the bursts of energy mediated by epinephrine and norepinephrine for long times. Instead, other hormones come into play. In a long-term stress response, the hypothalamus triggers the release of ACTH from the anterior pituitary gland. The adrenal cortex is stimulated by ACTH to release steroid hormones called **corticosteroids**. Corticosteroids turn on transcription of certain genes in the nuclei of target cells. They change enzyme concentrations in the cytoplasm and affect cellular metabolism. There are two main corticosteroids: glucocorticoids such as **cortisol**, and mineralocorticoids such as aldosterone. These hormones target the breakdown of fat into fatty acids in the adipose tissue. The fatty acids are released into the bloodstream for other tissues to use for ATP production. The **glucocorticoids** primarily affect glucose metabolism by stimulating glucose synthesis. Glucocorticoids also have anti-inflammatory properties through inhibition of the immune system. For example, cortisone is used as an anti-inflammatory medication; however, it cannot be used long term as it increases susceptibility to disease due to its immune-suppressing effects.

Mineralocorticoids function to regulate ion and water balance of the body. The hormone aldosterone stimulates the reabsorption of water and sodium ions in the kidney, which results in increased blood pressure and volume.

Hypersecretion of glucocorticoids can cause a condition known as **Cushing's disease**, characterized by a shifting of fat storage areas of the body. This can cause the accumulation of adipose tissue in the face and neck, and excessive glucose in the blood. Hyposecretion of the corticosteroids can cause **Addison's disease**, which may result in bronzing of the skin, hypoglycemia, and low electrolyte levels in the blood.

Summary

Water levels in the body are controlled by antidiuretic hormone (ADH), which is produced in the hypothalamus and triggers the reabsorption of water by the kidneys. Underproduction of ADH can cause diabetes insipidus. Aldosterone, a hormone produced by the adrenal cortex of the kidneys,

enhances Na^+ reabsorption from the extracellular fluids and subsequent water reabsorption by diffusion. The renin-angiotensin-aldosterone system is one way that aldosterone release is controlled.

The reproductive system is controlled by the gonadotropins follicle-stimulating hormone (FSH) and luteinizing hormone (LH), which are produced by the pituitary gland. Gonadotropin release is controlled by the hypothalamic hormone gonadotropin-releasing hormone (GnRH). FSH stimulates the maturation of sperm cells in males and is inhibited by the hormone inhibin, while LH stimulates the production of the androgen testosterone. FSH stimulates egg maturation in females, while LH stimulates the production of estrogens and progesterone. *Estrogens* are a group of steroid hormones produced by the ovaries that trigger the development of secondary sex characteristics in females as well as control the maturation of the ova. In females, the pituitary also produces prolactin, which stimulates milk production after childbirth, and oxytocin, which stimulates uterine contraction during childbirth and milk let-down during suckling.

Insulin is produced by the pancreas in response to rising blood glucose levels and allows cells to utilize blood glucose and store excess glucose for later use. Diabetes mellitus is caused by reduced insulin activity and causes high blood glucose levels, or hyperglycemia. Glucagon is released by the pancreas in response to low blood glucose levels and stimulates the breakdown of glycogen into glucose, which can be used by the body. The body's basal metabolic rate is controlled by the thyroid hormones thyroxine (T_4) and triiodothyronine (T_3). The anterior pituitary produces thyroid stimulating hormone (TSH), which controls the release of T_3 and T_4 from the thyroid gland. Iodine is necessary in the production of thyroid hormone, and the lack of iodine can lead to a condition called goiter.

Parathyroid hormone (PTH) is produced by the parathyroid glands in response to low blood Ca^{2+} levels. The parafollicular cells of the thyroid produce calcitonin, which reduces blood Ca^{2+} levels. Growth hormone (GH) is produced by the anterior pituitary and controls the growth rate of muscle and bone. GH action is indirectly mediated by insulin-like growth factors (IGFs). Short-term stress causes the hypothalamus to trigger the adrenal medulla to release epinephrine and norepinephrine, which trigger the fight or flight response. Long-term stress causes the hypothalamus to trigger the anterior pituitary to release adrenocorticotropic hormone (ACTH), which causes the release of corticosteroids, glucocorticoids, and mineralocorticoids, from the adrenal cortex.

Exercises

1. Pancreatic tumors may cause excess secretion of glucagon. Type I diabetes results from the failure of the pancreas to produce insulin. Which of the following statement about these two conditions is true?
    1. A pancreatic tumor and type I diabetes will have the opposite effects on blood sugar levels.
    2. A pancreatic tumor and type I diabetes will both cause hyperglycemia.
    3. A pancreatic tumor and type I diabetes will both cause hypoglycemia.
    4. Both pancreatic tumors and type I diabetes result in the inability of cells to take up glucose.
2. Drinking alcoholic beverages causes an increase in urine output. This most likely occurs because

alcohol:

1. inhibits ADH release
2. stimulates ADH release
3. inhibits TSH release
4. stimulates TSH release

3. FSH and LH release from the anterior pituitary is stimulated by _____.

1. TSH
2. GnRH
3. T_3
4. PTH

4. What hormone is produced by beta cells of the pancreas?

1. T_3
2. glucagon
3. insulin
4. T_4

5. When blood calcium levels are low, PTH stimulates:

1. excretion of calcium from the kidneys
2. excretion of calcium from the intestinges
3. osteoblasts
4. osteoclasts

6. Name and describe a function of one hormone produced by the anterior pituitary and one hormone produced by the posterior pituitary.
7. Describe one direct action of growth hormone (GH).

Answers

1. B
2. A
3. B
4. C
5. D
6. In addition to producing FSH and LH, the anterior pituitary 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. The posterior pituitary releases the

hormone oxytocin, which stimulates 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 vagina stimulates oxytocin release in childbirth. Contractions increase in intensity as blood levels of oxytocin rise until the birth is complete.

7. Hormonal regulation is required for the growth and replication of most cells in the body. Growth hormone (GH), produced by the anterior pituitary, accelerates the rate of protein synthesis, particularly in skeletal muscles and bones. Growth hormone has direct and indirect mechanisms of action. The direct actions of GH include: 1) stimulation of fat breakdown (lipolysis) and release into the blood by adipocytes. This results in a switch by most tissues from utilizing glucose as an energy source to utilizing fatty acids. This process is called a glucose-sparing effect. 2) In the liver, GH stimulates glycogen breakdown, which is then released into the blood as glucose. Blood glucose levels increase as most tissues are utilizing fatty acids instead of glucose for their energy needs. The GH mediated increase in blood glucose levels is called a diabetogenic effect because it is similar to the high blood glucose levels seen in diabetes mellitus.

Glossary

Addison's disease
disorder caused by the hyposecretion of corticosteroids

acromegaly
condition caused by overproduction of GH in adults

adrenocorticotropic hormone (ACTH)
hormone released by the anterior pituitary, which stimulates the adrenal cortex to release corticosteroids during the long-term stress response

aldosterone
steroid hormone produced by the adrenal cortex that stimulates the reabsorption of Na^+ from extracellular fluids and secretion of K^+.

androgen
male sex hormone such as testosterone

antidiuretic hormone (ADH)
hormone produced by the hypothalamus and released by the posterior pituitary that increases water reabsorption by the kidneys

calcitonin
hormone produced by the parafollicular cells of the thyroid gland that functions to lower blood Ca^{2+} levels and promote bone growth

corticosteroid
hormone released by the adrenal cortex in response to long-term stress

cortisol
glucocorticoid produced in response to stress

Cushing's disease
disorder caused by the hypersecretion of glucocorticoids

diabetes insipidus
disorder caused by underproduction of ADH

diabetes mellitus
disorder caused by low levels of insulin activity

diabetogenic effect
effect of GH that causes blood glucose levels to rise similar to diabetes mellitus

epinephrine
hormone released by the adrenal medulla in response to a short term stress

follicle-stimulating hormone (FSH)
hormone produced by the anterior pituitary that stimulates gamete production

gigantism
condition caused by overproduction of GH in children

glucagon
hormone produced by the alpha cells of the pancreas in response to low blood sugar; functions to raise blood sugar levels

glucocorticoid
corticosteroid that affects glucose metabolism

gluconeogenesis
synthesis of glucose from amino acids

glucose-sparing effect
effect of GH that causes tissues to use fatty acids instead of glucose as an energy source

glycogenolysis
breakdown of glycogen into glucose

goiter
enlargement of the thyroid gland caused by insufficient dietary iodine levels

gonadotropin
hormone that regulates the gonads, including FSH and LH

growth hormone (GH)
hormone produced by the anterior pituitary that promotes protein synthesis and body growth

growth hormone-inhibiting hormone (GHIH)
hormone produced by the hypothalamus that inhibits growth hormone production, also called somatostatin

growth hormone-releasing hormone (GHRH)
hormone released by the hypothalamus that triggers the release of GH

hyperglycemia
high blood sugar level

hyperthyroidism
overactivity of the thyroid gland

hypoglycemia
low blood sugar level

hypothyroidism
underactivity of the thyroid gland

insulin-like growth factor (IGF)
growth-promoting protein produced by the liver

insulin
hormone produced by the beta cells of the pancreas in response to high blood glucose levels; functions to lower blood glucose levels

mineralocorticoid
corticosteroid that affects ion and water balance

norepinephrine
hormone released by the adrenal medulla in response to a short-term stress hormone production by the gonads

osmoreceptor
receptor in the hypothalamus that monitors the concentration of electrolytes in the blood

oxytocin
hormone released by the posterior pituitary to stimulate uterine contractions during childbirth and milk let-down in the mammary glands

parathyroid gland
gland located on the surface of the thyroid that produces parathyroid hormone

parathyroid hormone (PTH)
hormone produced by the parathyroid glands in response to low blood Ca^{2+} levels; functions to raise blood Ca^{2+} levels

pituitary dwarfism
condition caused by underproduction of GH in children

pituitary gland
endocrine gland located at the base of the brain composed of an anterior and posterior region; also called hypophysis

pituitary stalk
(also, infundibulum) stalk that connects the pituitary gland to the hypothalamus

prolactin (PRL)
hormone produced by the anterior pituitary that stimulates milk production

prolactin-inhibiting hormone
hormone produced by the hypothalamus that inhibits the release of prolactin

prolactin-releasing hormone
hormone produced by the hypothalamus that stimulates the release of prolactin

renin
enzyme produced by the juxtaglomerular apparatus of the kidneys that reacts with angiotensinogen to cause the release of aldosterone

thyroglobulin
glycoprotein found in the thyroid that is converted into thyroid hormone

thyroid gland
endocrine gland located in the neck that produces thyroid hormones thyroxine and triiodothyronine

thyroid-stimulating hormone (TSH)
hormone produced by the anterior pituitary that controls the release of T_3 and T_4 from the thyroid gland

thyroxine (tetraiodothyronine, T_4)
thyroid hormone that controls the basal metabolic rate

triiodothyronine (T_3)
thyroid hormone that controls the basal metabolic rate

18.4 Regulation of Hormone Production

Learning Objectives

By the end of this section, you will be able to:

- Explain how hormone production is regulated
- Discuss the different stimuli that control hormone levels in the body

Hormone production and release are primarily controlled by negative feedback. In negative feedback systems, a stimulus elicits the release of a substance; once the substance reaches a certain level, it sends a signal that stops further release of the substance. In this way, the concentration of hormones in blood is maintained within a narrow range. For example, the anterior pituitary signals the thyroid to release thyroid hormones. Increasing levels of these hormones in the blood then give feedback to the hypothalamus and anterior pituitary to inhibit further signaling to the thyroid gland, as illustrated in Figure 18.14. There are three mechanisms by which endocrine glands are stimulated to synthesize and release hormones: humoral stimuli, hormonal stimuli, and neural stimuli.

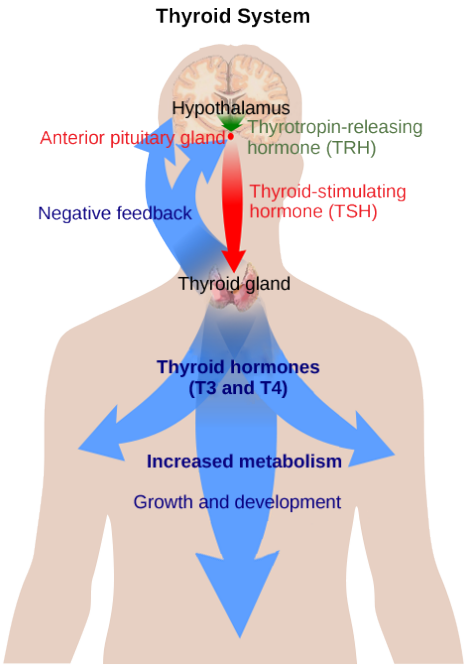


Figure 18.14.
The anterior pituitary stimulates the thyroid gland to release thyroid hormones T3 and T4. Increasing levels of these hormones in the blood results in feedback to the hypothalamus and anterior pituitary to inhibit further signaling to the thyroid gland. (credit: modification of work by Mikael Häggström)

Hyperthyroidism is a condition in which the thyroid gland is overactive. Hypothyroidism is a condition in which the thyroid gland is underactive. Which of the conditions are the following two patients most likely to have?

Patient A has symptoms including weight gain, cold sensitivity, low heart rate and fatigue.

Patient B has symptoms including weight loss, profuse sweating, increased heart rate and difficulty sleeping.

Humoral Stimuli

The term “humoral” is derived from the term “humor,” which refers to bodily fluids such as blood. A **humoral stimulus** refers to the control of hormone release in response to changes in extracellular fluids such as blood or the ion concentration in the blood. For example, a rise in blood glucose levels triggers the pancreatic release of insulin. Insulin causes blood glucose levels to drop, which signals the pancreas to stop producing insulin in a negative feedback loop.

Hormonal Stimuli

Hormonal stimuli refers to the release of a hormone in response to another hormone. A number of endocrine glands release hormones when stimulated by hormones released by other endocrine glands. For example, the hypothalamus produces hormones that stimulate the anterior portion of the pituitary gland. The anterior pituitary in turn releases hormones that regulate hormone production by other endocrine glands. The anterior pituitary releases the thyroid-stimulating hormone, which then stimulates the thyroid gland to produce the hormones T_3 and T_4. As blood concentrations of T_3 and T_4 rise, they inhibit both the pituitary and the hypothalamus in a negative feedback loop.

Neural Stimuli

In some cases, the nervous system directly stimulates endocrine glands to release hormones, which is referred to as **neural stimuli**. Recall that in a short-term stress response, the hormones epinephrine and norepinephrine are important for providing the bursts of energy required for the body to respond. Here, neuronal signaling from the sympathetic nervous system directly stimulates the adrenal medulla to release the hormones epinephrine and norepinephrine in response to stress.

Summary

Hormone levels are primarily controlled through negative feedback, in which rising levels of a hormone inhibit its further release. The three mechanisms of hormonal release are humoral stimuli, hormonal stimuli, and neural stimuli. Humoral stimuli refers to the control of hormonal release in response to changes in extracellular fluid levels or ion levels. Hormonal stimuli refers to the release of hormones in response to hormones released by other endocrine glands. Neural stimuli refers to the release of hormones in response to neural stimulation.

Exercises

1. Hyperthyroidism is a condition in which the thyroid gland is overactive. Hypothyroidism is a condition in which the thyroid gland is underactive. Which of the conditions are the following two patients most likely to have?
    1. Patient A has symptoms including weight gain, cold sensitivity, low heart rate and fatigue.
    2. Patient B has symptoms including weight loss, profuse sweating, increased heart rate and difficulty sleeping.
2. A rise in blood glucose levels triggers release of insulin from the pancreas. This mechanism of hormone production is stimulated by:
    1. humoral stimuli
    2. hormonal stimuli
    3. neural stimuli

4. negative stimuli

3. Which mechanism of hormonal stimulation would be affected if signaling and hormone release from the hypothalamus was blocked?

    1. humoral and hormonal stimuli
    2. hormonal and neural stimuli
    3. neural and humoral stimuli
    4. hormonal and negative stimuli

4. How is hormone production and release primarily controlled?
5. Compare and contrast hormonal and humoral stimuli.

Answers

1. Patient A has symptoms associated with decreased metabolism, and may be suffering from hypothyroidism. Patient B has symptoms associated with increased metabolism, and may be suffering from hyperthyroidism.
2. A
3. B
4. Hormone production and release are primarily controlled by negative feedback. In negative feedback systems, a stimulus causes the release of a substance whose effects then inhibit further release. In this way, the concentration of hormones in blood is maintained within a narrow range. For example, the anterior pituitary signals the thyroid to release thyroid hormones. Increasing levels of these hormones in the blood then feed back to the hypothalamus and anterior pituitary to inhibit further signaling to the thyroid gland.
5. The term humoral is derived from the term humor, which refers to bodily fluids such as blood. Humoral stimuli refer to the control of hormone release in response to changes in extracellular fluids such as blood or the ion concentration in the blood. For example, a rise in blood glucose levels triggers the pancreatic release of insulin. Insulin causes blood glucose levels to drop, which signals the pancreas to stop producing insulin in a negative feedback loop. Hormonal stimuli refer to the release of a hormone in response to another hormone. A number of endocrine glands release hormones when stimulated by hormones released by other endocrine organs. For example, the hypothalamus produces hormones that stimulate the anterior pituitary. The anterior pituitary in turn releases hormones that regulate hormone production by other endocrine glands. For example, the anterior pituitary releases thyroid-stimulating hormone, which stimulates the thyroid gland to produce the hormones T_3 and T_4. As blood concentrations of T_3 and T_4 rise they inhibit both the pituitary and the hypothalamus in a negative feedback loop.

Glossary

hormonal stimuli
release of a hormone in response to another hormone

humoral stimuli
control of hormone release in response to changes in extracellular fluids such as blood or the ion concentration in the blood

neural stimuli
stimulation of endocrine glands by the nervous system

18.5 Endocrine Glands

Learning Objectives

By the end of this section, you will be able to:

- Describe the role of different glands in the endocrine system
- Explain how the different glands work together to maintain homeostasis

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 18.15. 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.

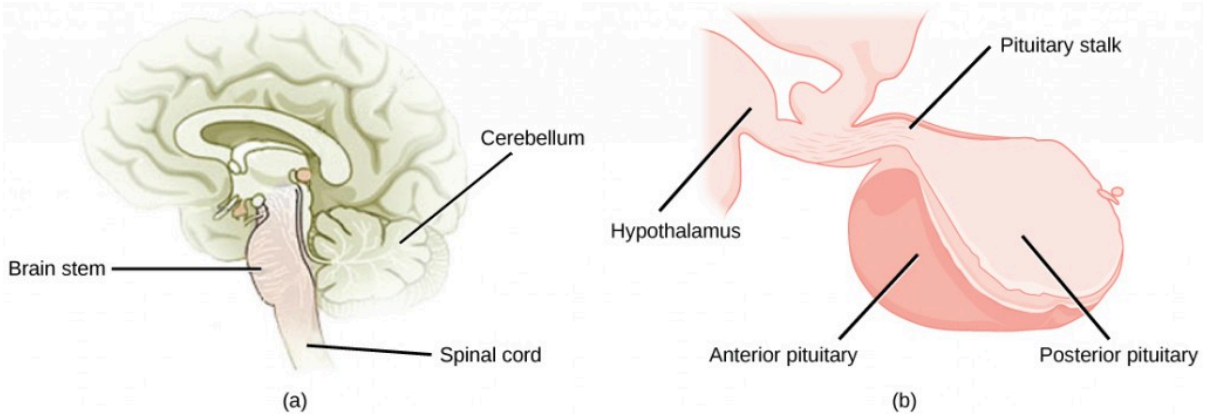


Figure 18.15.
The pituitary gland is located at (a) the base of the brain and (b) connected to the hypothalamus by the pituitary stalk. (credit a: modification of work by NCI; credit b: modification of work by Gray's Anatomy)

Anterior Pituitary

The **anterior pituitary** gland, or adenohypophysis, is surrounded by a capillary network that extends from the hypothalamus, down along the infundibulum, and to the anterior pituitary. This capillary network is a part of the **hypophyseal portal system** that carries substances from the hypothalamus to the anterior pituitary and hormones from the anterior pituitary into the circulatory system. A portal system carries blood from one capillary network to another; therefore, the hypophyseal portal system allows hormones produced by the hypothalamus to be carried directly to the anterior pituitary without first entering the circulatory system.

The anterior pituitary produces seven hormones: growth hormone (GH), prolactin (PRL), thyroid-stimulating hormone (TSH), melanin-stimulating hormone (MSH), adrenocorticotropic hormone (ACTH), follicle-stimulating hormone (FSH), and luteinizing hormone (LH). Anterior pituitary hormones are sometimes referred to as tropic hormones, because they control the functioning of other organs. While these hormones are produced by the anterior pituitary, their production is controlled by regulatory hormones produced by the hypothalamus. These regulatory hormones can be releasing hormones or inhibiting hormones, causing more or less of the anterior pituitary hormones to be secreted. These travel from the hypothalamus through the hypophyseal portal system to the anterior pituitary where they exert their effect. Negative feedback then regulates how much of these regulatory hormones are released and how much anterior pituitary hormone is secreted.

Posterior Pituitary

The **posterior pituitary** is significantly different in structure from the anterior pituitary. It is a part of the brain, extending down from the hypothalamus, and contains mostly nerve fibers and neuroglial cells, which support axons that extend from the hypothalamus to the posterior pituitary. The posterior pituitary and the infundibulum together are referred to as the neurohypophysis.

The hormones antidiuretic hormone (ADH), also known as vasopressin, and oxytocin are produced by neurons in the hypothalamus and transported within these axons along the infundibulum to the posterior pituitary. They are released into the circulatory system via neural signaling from the

hypothalamus. These hormones are considered to be posterior pituitary hormones, even though they are produced by the hypothalamus, because that is where they are released into the circulatory system. The posterior pituitary itself does not produce hormones, but instead stores hormones produced by the hypothalamus and releases them into the blood stream.

Thyroid Gland

The **thyroid gland** *i*s located in the neck, just below the larynx and in front of the trachea, as shown in Figure 18.16. 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.

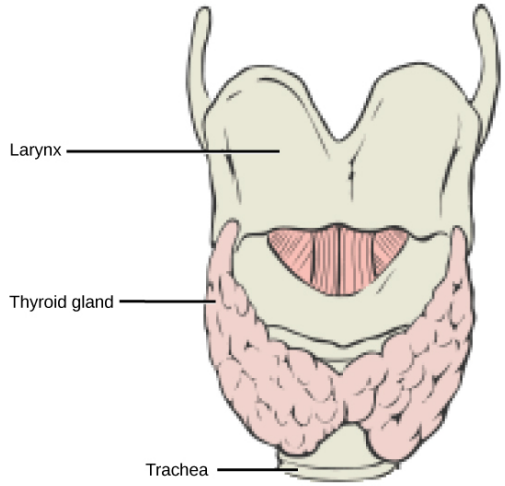


Figure 18.16.
This illustration shows the location of the thyroid gland.

The thyroid gland is made up of many spherical thyroid follicles, which are lined with a simple cuboidal epithelium. These follicles contain a viscous fluid, called **colloid**, which stores the glycoprotein thyroglobulin, the precursor to the thyroid hormones. The follicles produce hormones that can be stored in the colloid or released into the surrounding capillary network for transport to the rest of the body via the circulatory system.

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 18.17. 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.

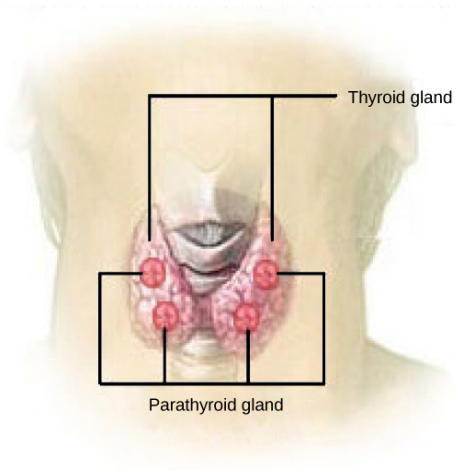


Figure 18.17.
The parathyroid glands are located on the posterior of the thyroid gland. (credit: modification of work by NCI)

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 is produced by chief cells of the parathyroid. PTH and calcitonin work in opposition to one another to maintain homeostatic Ca^{2+} levels in body fluids. Another type of cells, oxyphil cells, exist in the parathyroid but their function is not known. These hormones encourage bone growth, muscle mass, and blood cell formation in children and women.

Adrenal Glands

The **adrenal glands** are associated with the kidneys; one gland is located on top of each kidney as illustrated in Figure 18.18. The adrenal glands consist of an outer adrenal cortex and an inner adrenal medulla. These regions secrete different hormones.

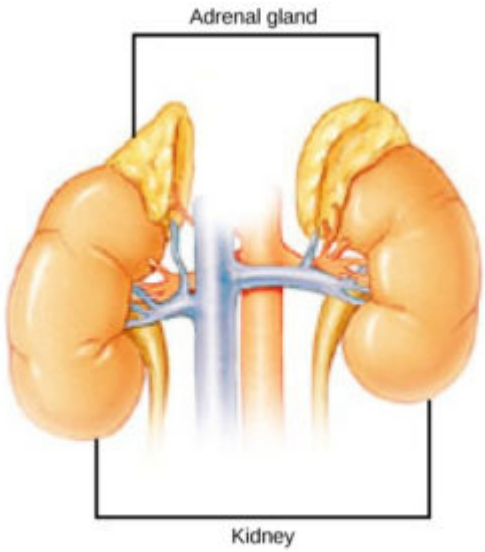


Figure 18.18. adrenal glands

Pancreas

The **pancreas**, illustrated in Figure 18.19, 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.

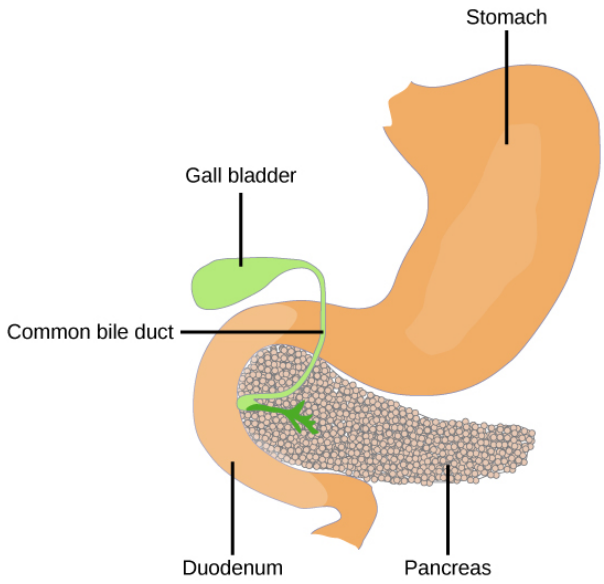


Figure 18.19. The pancreas is found underneath the stomach and points toward the spleen. (credit: modification of work by NCI)

The endocrine cells of the pancreas form clusters called pancreatic islets or the **islets of Langerhans**, as visible in the micrograph shown in Figure 18.20. 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.

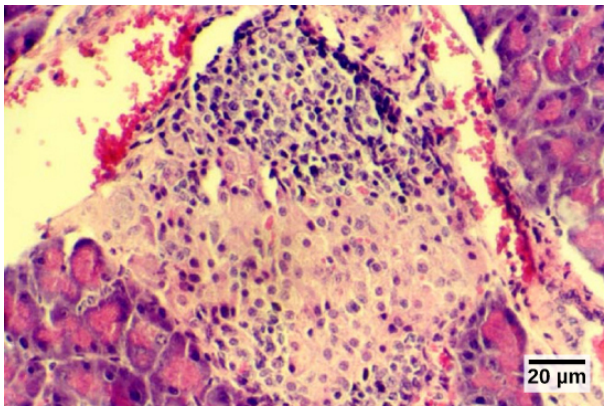


Figure 18.20.
The islets of Langerhans are clusters of endocrine cells found in the pancreas; they stain lighter than surrounding cells. (credit: modification of work by Muhammad T. Tabiin, Christopher P. White, Grant Morahan, and Bernard E. Tuch; scale-bar data from Matt Russell)

Pineal Gland

The pineal gland produces melatonin. The rate of melatonin production is affected by the photoperiod. Collaterals from the visual pathways innervate the pineal gland. During the day photoperiod, little melatonin is produced; however, melatonin production increases during the dark photoperiod (night). In some mammals, melatonin has an inhibitory affect on reproductive functions by decreasing production and maturation of sperm, oocytes, and reproductive organs. Melatonin is an effective antioxidant, protecting the CNS from free radicals such as nitric oxide and hydrogen peroxide. 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.

Table 18.1. 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
	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

Endocrine Gland	Associated Hormones	Effect
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

Organs with Secondary Endocrine Functions

There are several organs whose primary functions are non-endocrine but that also possess endocrine functions. These include the heart, kidneys, intestines, thymus, gonads, and adipose tissue.

The heart possesses endocrine cells in the walls of the atria that are specialized cardiac muscle cells. These cells release the hormone **atrial natriuretic peptide (ANP)** in response to increased blood volume. High blood volume causes the cells to be stretched, resulting in hormone release. ANP acts on the kidneys to reduce the reabsorption of Na^+, causing Na^+ and water to be excreted in the urine. ANP also reduces the amounts of renin released by the kidneys and aldosterone released by the adrenal cortex, further preventing the retention of water. In this way, ANP causes a reduction in blood volume and blood pressure, and reduces the concentration of Na^+ in the blood.

The gastrointestinal tract produces several hormones that aid in digestion. The endocrine cells are located in the mucosa of the GI tract throughout the stomach and small intestine. Some of the hormones produced include gastrin, secretin, and cholecystokinin, which are secreted in the presence of food, and some of which act on other organs such as the pancreas, gallbladder, and liver. They trigger the release of gastric juices, which help to break down and digest food in the GI tract.

While the adrenal glands associated with the kidneys are major **endocrine glands**, the kidneys themselves also possess endocrine function. Renin is released in response to decreased blood volume or pressure and is part of the renin-angiotensin-aldosterone system that leads to the release of aldosterone. Aldosterone then causes the retention of Na^+ and water, raising blood volume. The kidneys also release calcitriol, which aids in the absorption of Ca^{2+} and phosphate ions. **Erythropoietin (EPO)** is a protein hormone that triggers the formation of red blood cells in the bone marrow. EPO is released in response to low oxygen levels. Because red blood cells are oxygen carriers, increased production results in greater oxygen delivery throughout the body. EPO has been used by athletes to improve performance, as greater oxygen delivery to muscle cells allows for greater endurance. Because red blood cells increase the viscosity of blood, artificially high levels of EPO can cause severe health risks.

The **thymus** is found behind the sternum; it is most prominent in infants, becoming smaller in size through adulthood. The thymus produces hormones referred to as thymosins, which contribute to the development of the immune response.

Adipose tissue is a connective tissue found throughout the body. It produces the hormone **leptin** in

response to food intake. Leptin increases the activity of anorexigenic neurons and decreases that of orexigenic neurons, producing a feeling of satiety after eating, thus affecting appetite and reducing the urge for further eating. Leptin is also associated with reproduction. It must be present for GnRH and gonadotropin synthesis to occur. Extremely thin females may enter puberty late; however, if adipose levels increase, more leptin will be produced, improving fertility.

Summary

The pituitary gland is located at the base of the brain and is attached to the hypothalamus by the infundibulum. The anterior pituitary receives products from the hypothalamus by the hypophyseal portal system and produces six hormones. The posterior pituitary is an extension of the brain and releases hormones (antidiuretic hormone and oxytocin) produced by the hypothalamus.

The thyroid gland is located in the neck and is composed of two lobes connected by the isthmus. The thyroid is made up of follicle cells that produce the hormones thyroxine and triiodothyronine. Parafollicular cells of the thyroid produce calcitonin. The parathyroid glands lie on the posterior surface of the thyroid gland and produce parathyroid hormone.

The adrenal glands are located on top of the kidneys and consist of the renal cortex and renal medulla. The adrenal cortex is the outer part of the adrenal gland and produces the corticosteroids, glucocorticoids, and mineralocorticoids. The adrenal medulla is the inner part of the adrenal gland and produces the catecholamines epinephrine and norepinephrine.

The pancreas lies in the abdomen between the stomach and the small intestine. Clusters of endocrine cells in the pancreas form the islets of Langerhans, which are composed of alpha cells that release glucagon and beta cells that release insulin.

Some organs possess endocrine activity as a secondary function but have another primary function. The heart produces the hormone atrial natriuretic peptide, which functions to reduce blood volume, pressure, and Na^+ concentration. The gastrointestinal tract produces various hormones that aid in digestion. The kidneys produce renin, calcitriol, and erythropoietin. Adipose tissue produces leptin, which promotes satiety signals in the brain.

Exercises

1. Which endocrine glands are associated with the kidneys?
    1. thyroid glands
    2. pituitary glands
    3. adrenal glands
    4. gonads
2. Which of the following hormones is not produced by the anterior pituitary?
    1. oxytocin

2. growth hormone
3. prolactin
4. thyroid-stimulating hormone

3. What does aldosterone regulate, and how is it stimulated?
4. The adrenal medulla contains two types of secretory cells, what are they and what are their functions?

Answers

1. C
2. A
3. The main mineralocorticoid is aldosterone, which regulates the concentration of ions in urine, sweat, 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 an increase in blood potassium levels.
4. 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 75–80 percent of its secretions. Epinephrine and norepinephrine increase heart rate, breathing rate, cardiac muscle contractions, 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. These neural impulses originate from the hypothalamus in response to stress to prepare the body for the fight-or-flight response.

Glossary

adrenal cortex
outer portion of adrenal glands that produces corticosteroids

adrenal gland
endocrine glands associated with the kidneys

adrenal medulla
inner portion of adrenal glands that produces epinephrine and norepinephrine

alpha cell
endocrine cell of the pancreatic islets that produces the hormone glucagon

anterior pituitary
portion of the pituitary gland that produces six hormones; also called adenohypophysis

atrial natriuretic peptide (ANP)
hormone produced by the heart to reduce blood volume, pressure, and Na^+ concentration

beta cell
endocrine cell of the pancreatic islets that produces the hormone insulin

colloid
fluid inside the thyroid gland that contains the glycoprotein thyroglobulin

endocrine gland
gland that secretes hormones into the surrounding interstitial fluid, which then diffuse into blood and are carried to various organs and tissues within the body

erythropoietin (EPO)
hormone produced by the kidneys to stimulate red blood cell production in the bone marrow

hypophyseal portal system
system of blood vessels that carries hormones from the hypothalamus to the anterior pituitary

islets of Langerhans (pancreatic islets)
endocrine cells of the pancreas

isthmus
tissue mass that connects the two lobes of the thyroid gland

leptin
hormone produced by adipose tissue that promotes feelings of satiety and reduces hunger

pancreas
organ located between the stomach and the small intestine that contains exocrine and endocrine cells

parafollicular cell
thyroid cell that produces the hormone calcitonin

parathyroid gland
gland located on the surface of the thyroid that produces parathyroid hormone

pituitary gland
endocrine gland located at the base of the brain composed of an anterior and posterior region; also called hypophysis

pituitary stalk
(also, infundibulum) stalk that connects the pituitary gland to the hypothalamus

posterior pituitary
extension of the brain that releases hormones produced by the hypothalamus; along with the infundibulum, it is also referred to as the neurohypophysis

thymus
gland located behind the sternum that produces thymosin hormones that contribute to the development of the immune system

thyroid gland
endocrine gland located in the neck that produces thyroid hormones thyroxine and triiodothyronine

Chapter 19. The Musculoskeletal System

Figure 19.1. Improvements in the design of prostheses have allowed for a wider range of activities in recipients. (credit: modification of work by Stuart Grout)

Introduction

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. In severe cases in which the joint has suffered extensive damage, joint replacement surgery may be needed.

Progress in the science of prosthesis design has resulted in the development of artificial joints, with joint replacement surgery in the hips and knees being the most common. Replacement joints for shoulders, elbows, and fingers are also available. Even with this progress, there is still room for improvement in the design of prostheses. The state-of-the-art prostheses have limited durability and therefore wear out quickly, particularly in young or active individuals. Current research is focused on the use of new materials, such as carbon fiber, that may make prostheses more durable.

19.1 Types of Skeletal Systems

Learning Objectives

By the end of this section, you will be able to:

- Discuss the different types of skeletal systems
- Explain the role of the human skeletal system
- Compare and contrast different 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 19.2).

Figure 19.2.
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 an 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 19.3). This skeleton type provides defence against predators, supports the body, and allows for movement through the contraction of attached muscles. As with vertebrates, muscles must cross a joint inside the exoskeleton. Shortening of the muscle 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 because the exoskeleton does not grow as the organism grows.

Figure 19.3.
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. An example of a primitive endoskeletal structure is the spicules of sponges. The bones of

vertebrates are composed of tissues, whereas sponges have no true tissues (Figure 19.4). Endoskeletons provide support for the body, protect internal organs, and allow for movement through contraction of muscles attached to the skeleton.

Figure 19.4.
The skeletons of humans and horses are examples of endoskeletons. (credit: Ross Murphy)

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).

Concept in Action

Visit the interactive body site to build a virtual skeleton: select “skeleton” and click through the activity to place each bone.

Human Axial Skeleton

The **axial skeleton** forms the central axis of the body and includes the bones of the skull, ossicles of the middle ear, hyoid bone of the throat, vertebral column, and the thoracic cage (ribcage) (Figure 19.5). The function of the axial skeleton is to provide support and protection for the brain, the spinal cord, and the organs in the ventral body cavity. It provides a surface for the attachment of muscles that move the head, neck, and trunk, performs respiratory movements, and stabilizes parts of the appendicular skeleton.

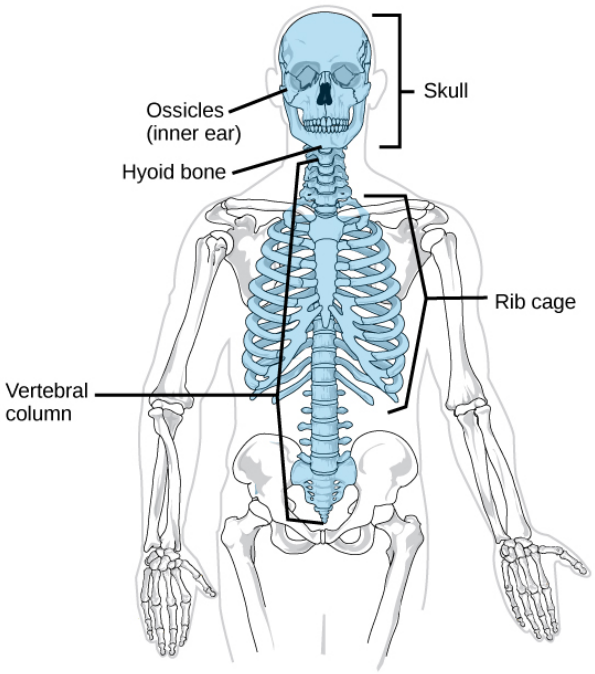


Figure 19.5.
The axial skeleton consists of the bones of the skull, ossicles of the middle ear, hyoid bone, vertebral column, and rib cage. (credit: modification of work by Mariana Ruiz Villareal)

The Skull

The bones of the **skull** support the structures of the face and protect the brain. The skull consists of 22 bones, which are divided into two categories: cranial bones and facial bones. The **cranial bones** are eight bones that form the cranial cavity, which encloses the brain and serves as an attachment site for the muscles of the head and neck. The eight cranial bones are the frontal bone, two parietal bones, two temporal bones, occipital bone, sphenoid bone, and the ethmoid bone. Although the bones developed separately in the embryo and fetus, in the adult, they are tightly fused with connective tissue and adjoining bones do not move (Figure 19.6).

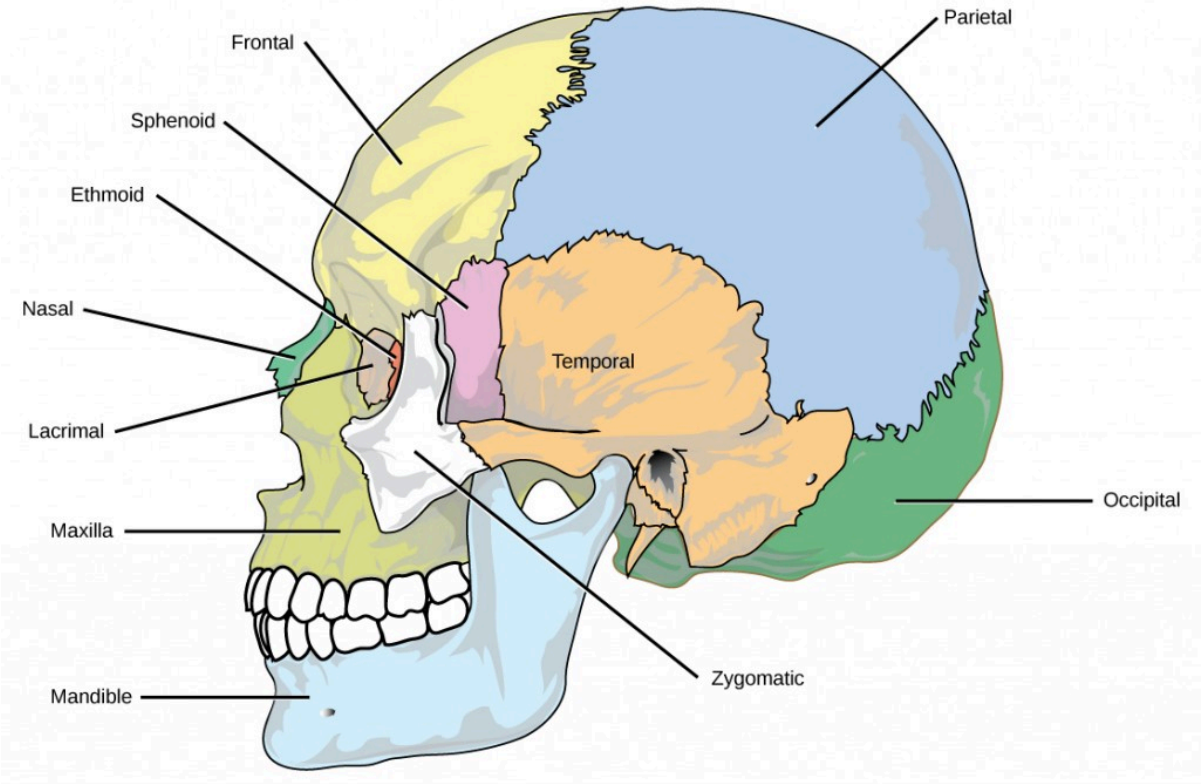


Figure 19.6.
The bones of the skull support the structures of the face and protect the brain. (credit: modification of work by Mariana Ruiz Villareal)

The **auditory ossicles** of the middle ear transmit sounds from the air as vibrations to the fluid-filled cochlea. The auditory ossicles consist of six bones: two malleus bones, two incus bones, and two stapes on each side. These are the smallest bones in the body and are unique to mammals.

Fourteen **facial bones** form the face, provide cavities for the sense organs (eyes, mouth, and nose), protect the entrances to the digestive and respiratory tracts, and serve as attachment points for facial muscles. The 14 facial bones are the nasal bones, the maxillary bones, zygomatic bones, palatine, vomer, lacrimal bones, the inferior nasal conchae, and the mandible. All of these bones occur in pairs except for the mandible and the vomer (Figure 19.7).

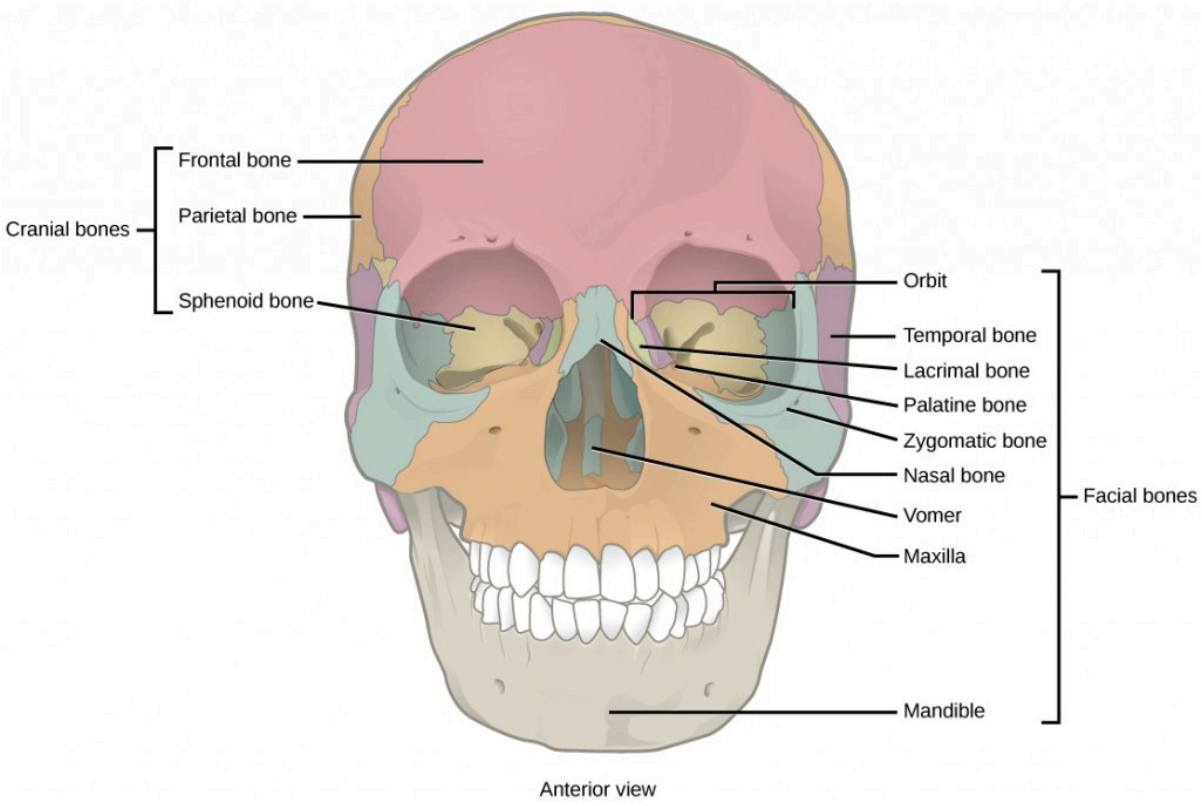


Figure 19.7.
The cranial bones, including the frontal, parietal, and sphenoid bones, cover the top of the head. The facial bones of the skull form the face and provide cavities for the eyes, nose, and mouth.

Although it is not found in the skull, the hyoid bone is considered a component of the axial skeleton. The *hyoid bone* lies below the mandible in the front of the neck. It acts as a movable base for the tongue and is connected to muscles of the jaw, larynx, and tongue. The mandible articulates with the base of the skull. The mandible controls the opening to the airway and gut. In animals with teeth, the mandible brings the surfaces of the teeth in contact with the maxillary teeth.

The Vertebral Column

The **vertebral column**, or spinal column, surrounds and protects the spinal cord, supports the head, and acts as an attachment point for the ribs and muscles of the back and neck. The adult vertebral column comprises 26 bones: the 24 vertebrae, the sacrum, and the coccyx bones. In the adult, the sacrum is typically composed of five vertebrae that fuse into one. The coccyx is typically 3–4 vertebrae that fuse into one. Around the age of 70, the sacrum and the coccyx may fuse together. We begin life with approximately 33 vertebrae, but as we grow, several vertebrae fuse together. The adult vertebrae are further divided into the 7 cervical vertebrae, 12 thoracic vertebrae, and 5 lumbar vertebrae (Figure 19.8).

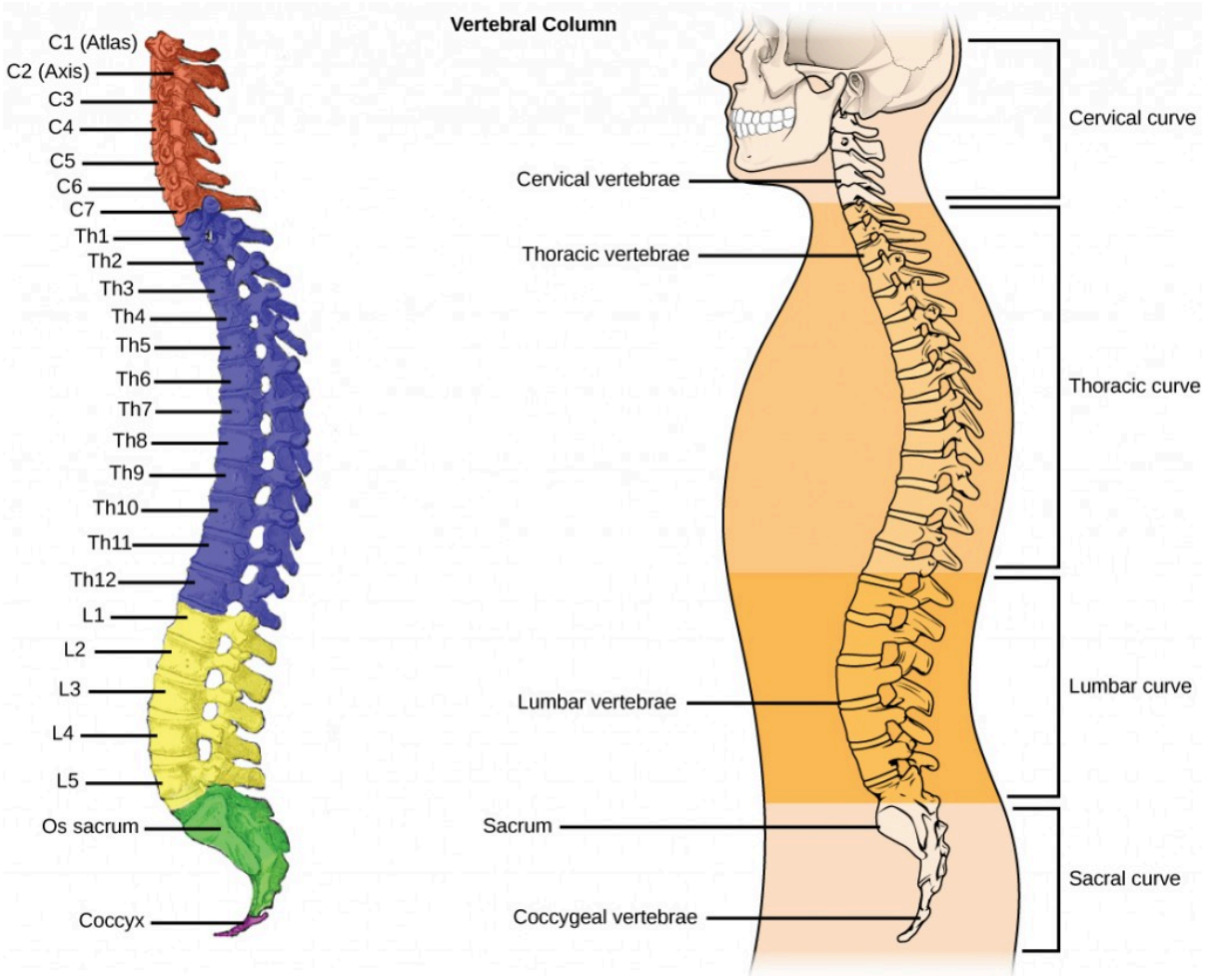


Figure 19.8. (a) The vertebral column consists of seven cervical vertebrae (C1–7) twelve thoracic vertebrae (Th1–12), five lumbar vertebrae (L1–5), the os sacrum, and the coccyx. (b) Spinal curves increase the strength and flexibility of the spine. (credit a: modification of work by Uwe Gille based on original work by Gray's Anatomy; credit b: modification of work by NCI, NIH)

Each vertebral body has a large hole in the center through which the nerves of the spinal cord pass. There is also a notch on each side through which the spinal nerves, which serve the body at that level, can exit from the spinal cord. The vertebral column is approximately 71 cm (28 inches) in adult male humans and is curved, which can be seen from a side view. The names of the spinal curves correspond to the region of the spine in which they occur. The thoracic and sacral curves are concave (curve inwards relative to the front of the body) and the cervical and lumbar curves are convex (curve outwards relative to the front of the body). The arched curvature of the vertebral column increases its strength and flexibility, allowing it to absorb shocks like a spring (Figure 19.8).

Intervertebral discs composed of fibrous cartilage lie between adjacent vertebral bodies from the second cervical vertebra to the sacrum. Each disc is part of a joint that allows for some movement of the spine and acts as a cushion to absorb shocks from movements such as walking and running. Intervertebral discs also act as ligaments to bind vertebrae together. The inner part of discs, the nucleus pulposus, hardens as people age and becomes less elastic. This loss of elasticity diminishes its ability to absorb shocks.

The Thoracic Cage

The **thoracic cage**, also known as the ribcage, is the skeleton of the chest, and consists of the ribs, sternum, thoracic vertebrae, and costal cartilages (Figure 19.9). The thoracic cage encloses and protects the organs of the thoracic cavity, including the heart and lungs. It also provides support for the shoulder girdles and upper limbs, and serves as the attachment point for the diaphragm, muscles of the back, chest, neck, and shoulders. Changes in the volume of the thorax enable breathing.

The **sternum**, or breastbone, is a long, flat bone located at the anterior of the chest. It is formed from three bones that fuse in the adult. The **ribs** are 12 pairs of long, curved bones that attach to the thoracic vertebrae and curve toward the front of the body, forming the ribcage. Costal cartilages connect the anterior ends of the ribs to the sternum, with the exception of rib pairs 11 and 12, which are free-floating ribs.

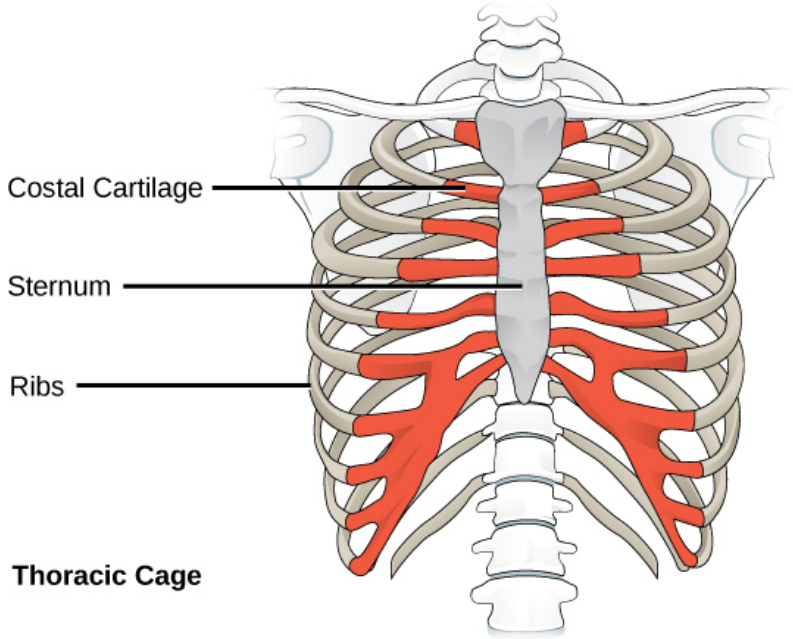


Figure 19.9.
The thoracic cage, or rib cage, protects the heart and the lungs.
(credit: modification of work by NCI, NIH)

Human Appendicular Skeleton

The **appendicular skeleton** is composed of the bones of the upper limbs (which function to grasp and manipulate objects) and the lower limbs (which permit locomotion). It also includes the pectoral girdle, or shoulder girdle, that attaches the upper limbs to the body, and the pelvic girdle that attaches the lower limbs to the body (Figure 19.10).

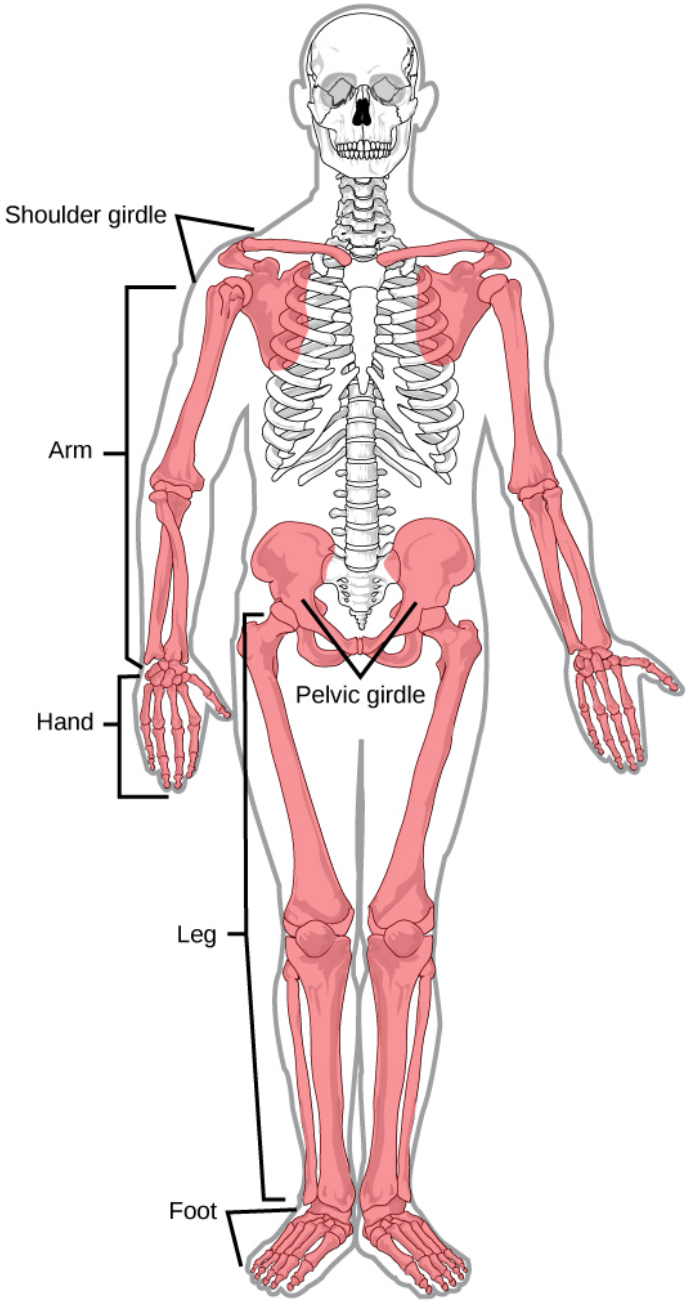


Figure 19.10. The appendicular skeleton is composed of the bones of the pectoral limbs (arm, forearm, hand), the pelvic limbs (thigh, leg, foot), the pectoral girdle, and the pelvic girdle. (credit: modification of work by Mariana Ruiz Villareal)

The Pectoral Girdle

The **pectoral girdle** bones provide the points of attachment of the upper limbs to the axial skeleton. The human pectoral girdle consists of the clavicle (or collarbone) in the anterior, and the scapula (or shoulder blades) in the posterior (Figure 19.11).

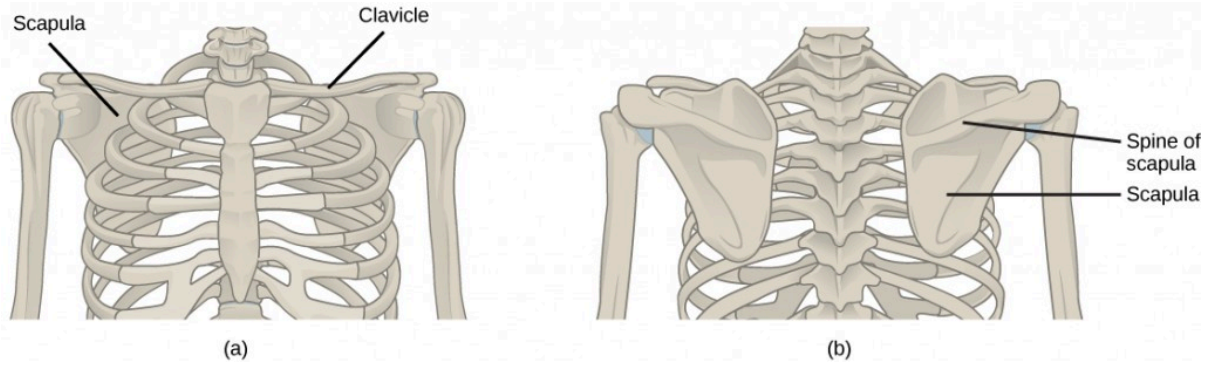


Figure 19.11.
(a) The pectoral girdle in primates consists of the clavicles and scapulae. (b) The posterior view reveals the spine of the scapula to which muscle attaches.

The **clavicles** are S-shaped bones that position the arms on the body. The clavicles lie horizontally across the front of the thorax (chest) just above the first rib. These bones are fairly fragile and are susceptible to fractures. For example, a fall with the arms outstretched causes the force to be transmitted to the clavicles, which can break if the force is excessive. The clavicle articulates with the sternum and the scapula.

The **scapulae** are flat, triangular bones that are located at the back of the pectoral girdle. They support the muscles crossing the shoulder joint. A ridge, called the spine, runs across the back of the scapula and can easily be felt through the skin (Figure 19.11). The spine of the scapula is a good example of a bony protrusion that facilitates a broad area of attachment for muscles to bone.

The Upper Limb

The upper limb contains 30 bones in three regions: the arm (shoulder to elbow), the forearm (ulna and radius), and the wrist and hand (Figure 19.12).

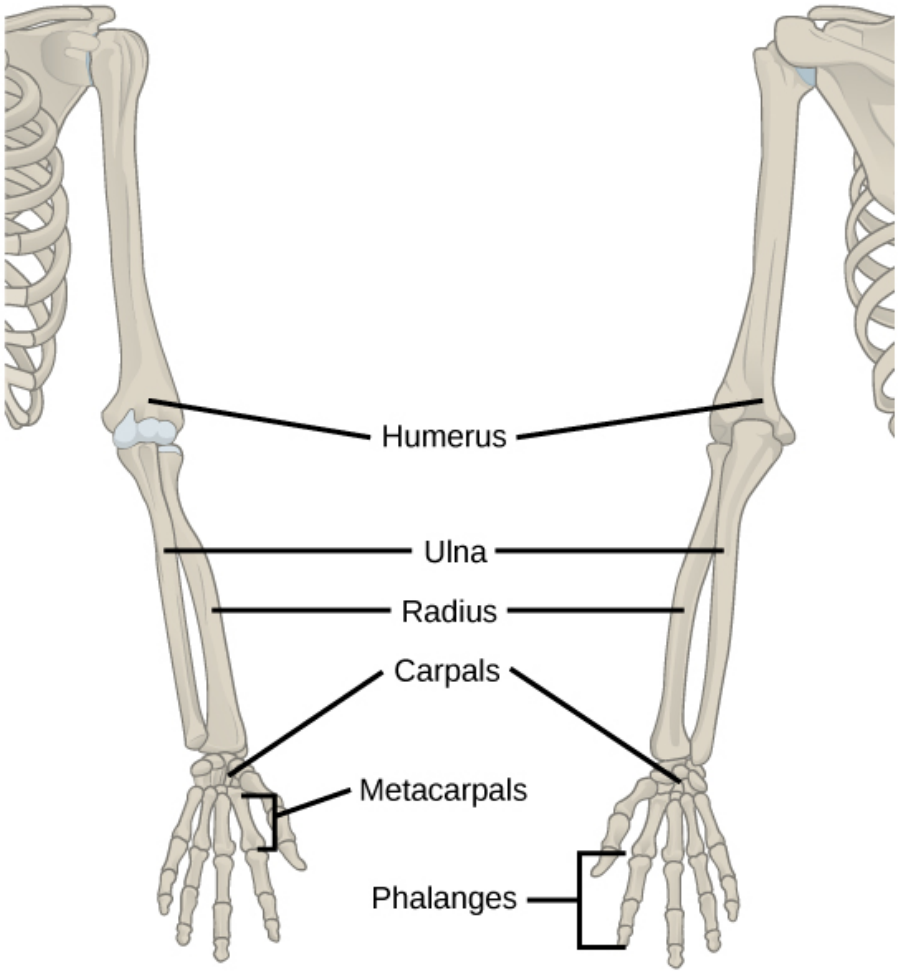


Figure 19.12.
The upper limb consists of the humerus of the upper arm, the radius and ulna of the forearm, eight bones of the carpus, five bones of the metacarpus, and 14 bones of the phalanges.

An **articulation** is any place at which two bones are joined. The **humerus** is the largest and longest bone of the upper limb and the only bone of the arm. It articulates with the scapula at the shoulder and with the forearm at the elbow. The **forearm** extends from the elbow to the wrist and consists of two bones: the ulna and the radius. The **radius** is located along the lateral (thumb) side of the forearm and articulates with the humerus at the elbow. The **ulna** is located on the medial aspect (pinky-finger side) of the forearm. It is longer than the radius. The ulna articulates with the humerus at the elbow. The radius and ulna also articulate with the carpal bones and with each other, which in vertebrates enables a variable degree of rotation of the carpus with respect to the long axis of the limb. The hand includes the eight bones of the **carpus** (wrist), the five bones of the **metacarpus** (palm), and the 14 bones of the **phalanges** (digits). Each digit consists of three phalanges, except for the thumb, when present, which has only two.

The Pelvic Girdle

The **pelvic girdle** attaches to the lower limbs of the axial skeleton. Because it is responsible for bearing

the weight of the body and for locomotion, the pelvic girdle is securely attached to the axial skeleton by strong ligaments. It also has deep sockets with robust ligaments to securely attach the femur to the body. The pelvic girdle is further strengthened by two large hip bones. In adults, the hip bones, or **coxal bones** are formed by the fusion of three pairs of bones: the ilium, ischium, and pubis. The pelvis joins together in the anterior of the body at a joint called the pubic symphysis and with the bones of the sacrum at the posterior of the body.

The female pelvis is slightly different from the male pelvis. Over generations of evolution, females with a wider pubic angle and larger diameter pelvic canal reproduced more successfully. Therefore, their offspring also had pelvic anatomy that enabled successful childbirth (Figure 19.13).

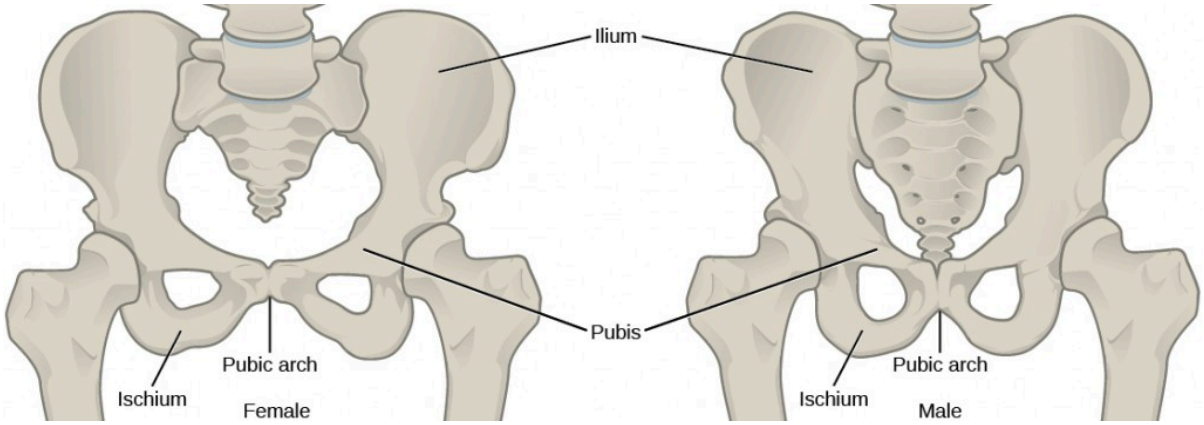


Figure 19.13.
To adapt to reproductive fitness, the (a) female pelvis is lighter, wider, shallower, and has a broader angle between the pubic bones than (b) the male pelvis.

The Lower Limb

The **lower limb** consists of the thigh, the leg, and the foot. The bones of the lower limb are the femur (thigh bone), patella (kneecap), tibia and fibula (bones of the leg), tarsals (bones of the ankle), and metatarsals and phalanges (bones of the foot) (Figure 19.14). The bones of the lower limbs are thicker and stronger than the bones of the upper limbs because of the need to support the entire weight of the body and the resulting forces from locomotion. In addition to evolutionary fitness, the bones of an individual will respond to forces exerted upon them.

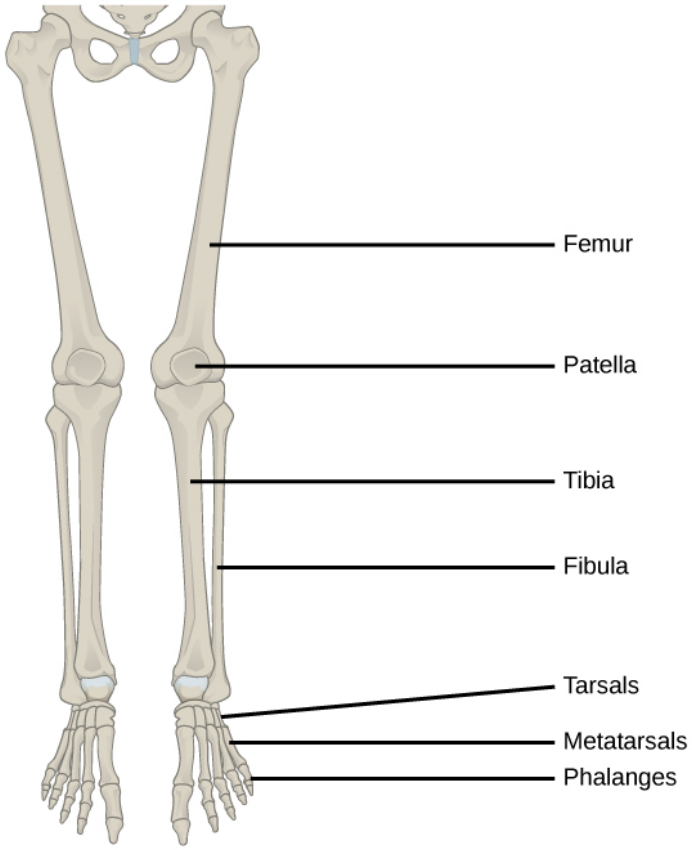


Figure 19.14.
The lower limb consists of the thigh (femur), kneecap (patella), leg (tibia and fibula), ankle (tarsals), and foot (metatarsals and phalanges) bones.

The **femur**, or thighbone, is the longest, heaviest, and strongest bone in the body. The femur and pelvis form the hip joint at the proximal end. At the distal end, the femur, tibia, and patella form the knee joint. The **patella**, or kneecap, is a triangular bone that lies anterior to the knee joint. The patella is embedded in the tendon of the femoral extensors (quadriceps). It improves knee extension by reducing friction. The **tibia**, or shinbone, is a large bone of the leg that is located directly below the knee. The tibia articulates with the femur at its proximal end, with the fibula and the tarsal bones at its distal end. It is the second largest bone in the human body and is responsible for transmitting the weight of the body from the femur to the foot. The **fibula**, or calf bone, parallels and articulates with the tibia. It does not articulate with the femur and does not bear weight. The fibula acts as a site for muscle attachment and forms the lateral part of the ankle joint.

The **tarsals** are the seven bones of the ankle. The ankle transmits the weight of the body from the tibia and the fibula to the foot. The **metatarsals** are the five bones of the foot. The phalanges are the 14 bones of the toes. Each toe consists of three phalanges, except for the big toe that has only two (Figure 19.15). Variations exist in other species; for example, the horse’s metacarpals and metatarsals are oriented vertically and do not make contact with the substrate.

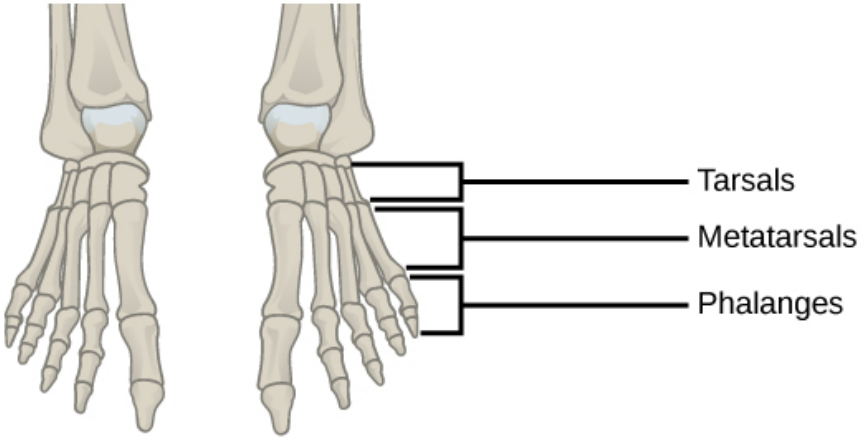


Figure 19.15.
This drawing shows the bones of the human foot and ankle, including the metatarsals and the phalanges.

Evolution of Body Design for Locomotion on Land

The transition of vertebrates onto land required a number of changes in body design, as movement on land presents a number of challenges for animals that are adapted to movement in water. The buoyancy of water provides a certain amount of lift, and a common form of movement by fish is lateral undulations of the entire body. This back and forth movement pushes the body against the water, creating forward movement. In most fish, the muscles of paired fins attach to girdles within the body, allowing for some control of locomotion. As certain fish began moving onto land, they retained their lateral undulation form of locomotion (anguilliform). However, instead of pushing against water, their fins or flippers became points of contact with the ground, around which they rotated their bodies.

The effect of gravity and the lack of buoyancy on land meant that body weight was suspended on the limbs, leading to increased strengthening and ossification of the limbs. The effect of gravity also required changes to the axial skeleton. Lateral undulations of land animal vertebral columns cause torsional strain. A firmer, more ossified vertebral column became common in terrestrial tetrapods because it reduces strain while providing the strength needed to support the body’s weight. In later tetrapods, the vertebrae began allowing for vertical motion rather than lateral flexion. Another change in the axial skeleton was the loss of a direct attachment between the pectoral girdle and the head. This reduced the jarring to the head caused by the impact of the limbs on the ground. The vertebrae of the neck also evolved to allow movement of the head independently of the body.

The appendicular skeleton of land animals is also different from aquatic animals. The shoulders attach to the pectoral girdle through muscles and connective tissue, thus reducing the jarring of the skull. Because of a lateral undulating vertebral column, in early tetrapods, the limbs were splayed out to the side and movement occurred by performing “push-ups.” The vertebrae of these animals had to move side-to-side in a similar manner to fish and reptiles. This type of motion requires large muscles to move the limbs toward the midline; it was almost like walking while doing push-ups, and it is not an efficient use of energy. Later tetrapods have their limbs placed under their bodies, so that each stride requires less force to move forward. This resulted in decreased adductor muscle size and an increased range of motion of the scapulae. This also restricts movement primarily to one plane, creating forward motion rather than moving the limbs upward as well as forward. The femur and humerus were also rotated, so

that the ends of the limbs and digits were pointed forward, in the direction of motion, rather than out to the side. By placement underneath the body, limbs can swing forward like a pendulum to produce a stride that is more efficient for moving over land.

Summary

The three types of skeleton designs are hydrostatic skeletons, exoskeletons, and endoskeletons. A hydrostatic skeleton is formed by a fluid-filled compartment held under hydrostatic pressure; movement is created by the muscles producing pressure on the fluid. An exoskeleton is a hard external skeleton that protects the outer surface of an organism and enables movement through muscles attached on the inside. An endoskeleton is an internal skeleton composed of hard, mineralized tissue that also enables movement by attachment to muscles. The human skeleton is an endoskeleton that is composed of the axial and appendicular skeleton. The axial skeleton is composed of the bones of the skull, ossicles of the ear, hyoid bone, vertebral column, and ribcage. The skull consists of eight cranial bones and 14 facial bones. Six bones make up the ossicles of the middle ear, while the hyoid bone is located in the neck under the mandible. The vertebral column contains 26 bones, and it surrounds and protects the spinal cord. The thoracic cage consists of the sternum, ribs, thoracic vertebrae, and costal cartilages. The appendicular skeleton is made up of the limbs of the upper and lower limbs. The pectoral girdle is composed of the clavicles and the scapulae. The upper limb contains 30 bones in the arm, the forearm, and the hand. The pelvic girdle attaches the lower limbs to the axial skeleton. The lower limb includes the bones of the thigh, the leg, and the foot.

Exercises

1. Which of the following statements about bone tissue is false?
    1. Compact bone tissue is made of cylindrical osteons that are aligned such that they travel the length of the bone.
    2. Haversian canals contain blood vessels only.
    3. Haversian canals contain blood vessels and nerve fibers.
    4. Spongy tissue is found on the interior of the bone, and compact bone tissue is found on the exterior.
2. The forearm consists of the:
    1. radius and ulna
    2. radius and humerus
    3. ulna and humerous
    4. humerus and carpus
3. The pectoral girdle consists of the:
    1. clavicle and sternum
    2. sternum and scapula

3. clavicle and scapula
4. clavicle and coccyx

4. All of the following are groups of vertebrae except _____, which is a curvature.

    1. thoracic
    2. cervical
    3. lumbar
    4. pelvic

5. Which of these is a facial bone?

    1. frontal
    2. occipital
    3. lacrimal
    4. temporal

6. What are the major differences between the male pelvis and female pelvis that permit childbirth in females?
7. What are the major differences between the pelvic girdle and the pectoral girdle that allow the pelvic girdle to bear the weight of the body?

Answers

1. B
2. A
3. C
4. D
5. C
6. The female pelvis is tilted forward and is wider, lighter, and shallower than the male pelvis. It is also has a pubic angle that is broader than the male pelvis.
7. The pelvic girdle is securely attached to the body by strong ligaments, unlike the pectoral girdle, which is sparingly attached to the ribcage. The sockets of the pelvic girdle are deep, allowing the femur to be more stable than the pectoral girdle, which has shallow sockets for the scapula. Most tetrapods have 75 percent of their weight on the front legs because the head and neck are so heavy; the advantage of the shoulder joint is more degrees of freedom in movement.

Glossary

abduction
when a bone moves away from the midline of the body

actin
globular contractile protein that interacts with myosin for muscle contraction

appendicular skeleton
composed of the bones of the upper limbs, which function to grasp and manipulate objects, and the lower limbs, which permit locomotion

articulation
any place where two bones are joined

auditory ossicle
(also, middle ear) transduces sounds from the air into vibrations in the fluid-filled cochlea

axial skeleton
forms the central axis of the body and includes the bones of the skull, the ossicles of the middle ear, the hyoid bone of the throat, the vertebral column, and the thoracic cage (ribcage)

bone remodeling
replacement of old bone tissue by new bone tissue

bone
(also, osseous tissue) connective tissue that constitutes the endoskeleton

carpus
eight bones that comprise the wrist

clavicle
S-shaped bone that positions the arms laterally

compact bone
forms the hard external layer of all bones

coxal bone
hip bone

cranial bone
one of eight bones that form the cranial cavity that encloses the brain and serves as an attachment site for the muscles of the head and neck

diaphysis
central shaft of bone, contains bone marrow in a marrow cavity

endoskeleton
skeleton of living cells that produce a hard, mineralized tissue located within the soft tissue of organisms

epiphyseal plate
region between the diaphysis and epiphysis that is responsible for the lengthwise growth of long bones

epiphysis
rounded end of bone, covered with articular cartilage and filled with red bone marrow, which produces blood cells

exoskeleton
a secreted cellular product external skeleton that consists of a hard encasement on the surface of an organism

extension
movement in which the angle between the bones of a joint increases; opposite of flexion

facial bone
one of the 14 bones that form the face; provides cavities for the sense organs (eyes, mouth, and nose) and attachment points for facial muscles

femur
(also, thighbone) longest, heaviest, and strongest bone in the body

fibula
(also, calf bone) parallels and articulates with the tibia

flat bone
thin and relatively broad bone found where extensive protection of organs is required or where broad surfaces of muscle attachment are required

flexion
movement in which the angle between the bones decreases; opposite of extension

forearm
extends from the elbow to the wrist and consists of two bones: the ulna and the radius

Haversian canal
contains the bone’s blood vessels and nerve fibers

humerus
only bone of the arm

hydrostatic skeleton
skeleton that consists of aqueous fluid held under pressure in a closed body compartment

hyoid bone
lies below the mandible in the front of the neck

joint
point at which two or more bones meet

lamella
layer of compact tissue that surrounds a central canal called the Haversian canal

long bone
bone that is longer than wide, and has a shaft and two ends

lower limb
consists of the thigh, the leg, and the foot

metacarpus
five bones that comprise the palm

metatarsal
one of the five bones of the foot

myofibril
long cylindrical structures that lie parallel to the muscle fiber

myosin
contractile protein that interacts with actin for muscle contraction

osseous tissue
connective tissue that constitutes the endoskeleton

ossification
(also, osteogenesis) process of bone formation by osteoblasts

osteoblast
bone cell responsible for bone formation

osteoclast
large bone cells with up to 50 nuclei, responsible for bone remodeling

osteon
cylindrical structure aligned parallel to the long axis of the bone

patella
(also, kneecap) triangular bone that lies anterior to the knee joint

pectoral girdle
bones that transmit the force generated by the upper limbs to the axial skeleton

pelvic girdle
bones that transmit the force generated by the lower limbs to the axial skeleton

phalange
one of the bones of the fingers or toes

protraction
anterior movement of a bone in the horizontal plane

radius
bone located along the lateral (thumb) side of the forearm; articulates with the humerus at the elbow

rib
one of 12 pairs of long, curved bones that attach to the thoracic vertebrae and curve toward the front of the body to form the ribcage

scapula
flat, triangular bone located at the posterior pectoral girdle

skull
bone that supports the structures of the face and protects the brain

sternum
(also, breastbone) long, flat bone located at the front of the chest

suture
short fiber of connective tissue that holds the skull bones tightly in place; found only in the skull

symphysis
hyaline cartilage covers the end of the bone, but the connection between bones occurs through fibrocartilage; symphyses are found at the joints between vertebrae

tarsal
one of the seven bones of the ankle

thoracic cage
(also, ribcage) skeleton of the chest, which consists of the ribs, thoracic vertebrae, sternum, and costal cartilages

tibia
(also, shinbone) large bone of the leg that is located directly below the knee

tropomyosin
acts to block myosin binding sites on actin molecules, preventing cross-bridge formation and preventing contraction until a muscle receives a neuron signal

ulna
bone located on the medial aspect (pinky-finger side) of the forearm

vertebral column
(also, spine) surrounds and protects the spinal cord, supports the head, and acts as an attachment point for ribs and muscles of the back and neck

19.2 Bone

Learning Objectives

By the end of this section, you will be able to:

- Classify the different types of bones in the skeleton
- Explain the role of the different cell types in bone
- Explain how bone forms during development

Bone, or **osseous tissue**, is a connective tissue that constitutes the endoskeleton. It contains specialized cells and a matrix of mineral salts and collagen fibers.

The mineral salts primarily include hydroxyapatite, a mineral formed from calcium phosphate. **Calcification** is the process of deposition of mineral salts on the collagen fiber matrix that crystallizes and hardens the tissue. The process of calcification only occurs in the presence of collagen fibers.

The bones of the human skeleton are classified by their shape: long bones, short bones, flat bones, sutural bones, sesamoid bones, and irregular bones (Figure 19.16).

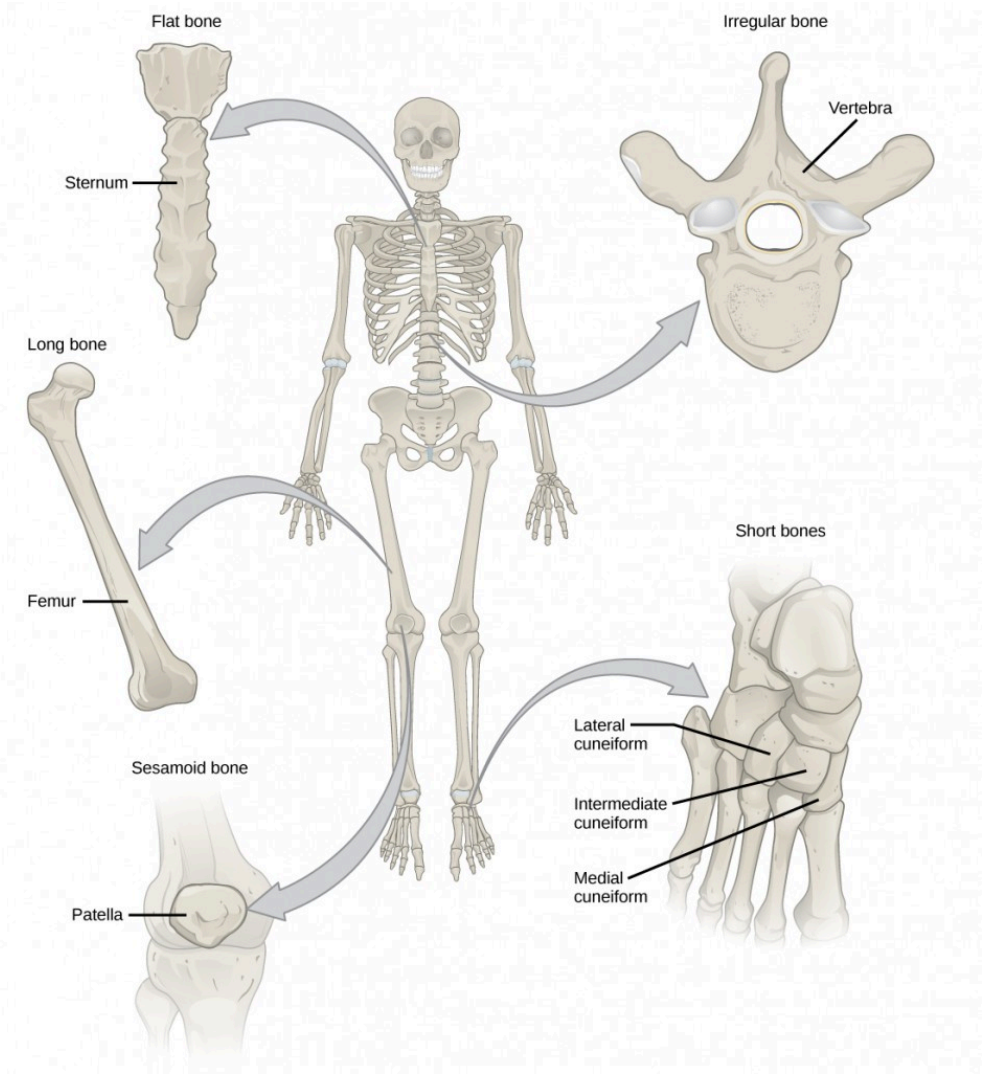


Figure 19.16. Shown are different types of bones: flat, irregular, long, short, and sesamoid.

Long bones are longer than they are wide and have a shaft and two ends. The **diaphysis**, or central shaft, contains bone marrow in a marrow cavity. The rounded ends, the **epiphyses**, are covered with articular cartilage and are filled with red bone marrow, which produces blood cells (Figure 19.17). Most of the limb bones are long bones—for example, the femur, tibia, ulna, and radius. Exceptions to this include the patella and the bones of the wrist and ankle.

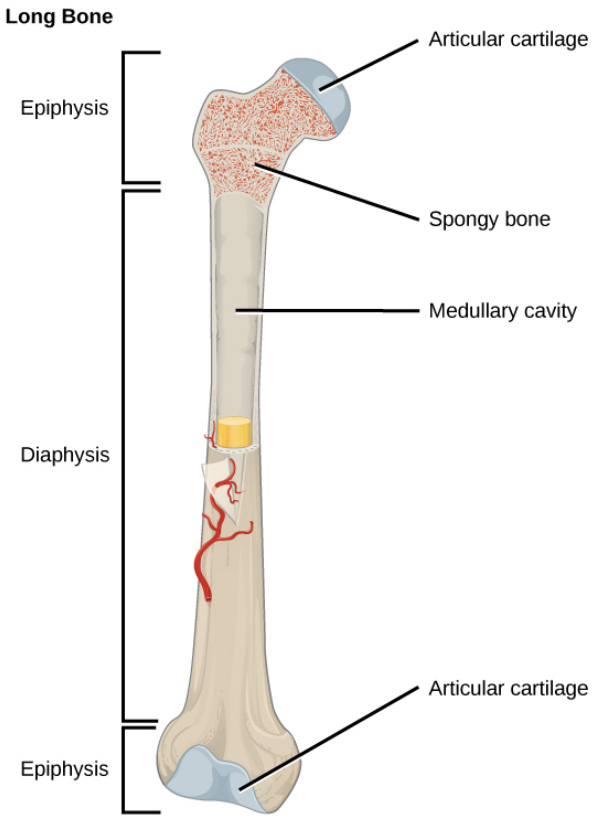


Figure 19.17.
The long bone is covered by articular cartilage at either end and contains bone marrow (shown in yellow in this illustration) in the marrow cavity.

Short bones, or cuboidal bones, are bones that are the same width and length, giving them a cube-like shape. For example, the bones of the wrist (carpals) and ankle (tarsals) are short bones (Figure 19.16).

Flat bones are thin and relatively broad bones that are found where extensive protection of organs is required or where broad surfaces of muscle attachment are required. Examples of flat bones are the sternum (breast bone), ribs, scapulae (shoulder blades), and the roof of the skull (Figure 19.16).

Irregular bones are bones with complex shapes. These bones may have short, flat, notched, or ridged surfaces. Examples of irregular bones are the vertebrae, hip bones, and several skull bones.

Sesamoid bones are small, flat bones and are shaped similarly to a sesame seed. The patellae are sesamoid bones (Figure 19.18). Sesamoid bones develop inside tendons and may be found near joints at the knees, hands, and feet.

Figure 19.18.
The patella of the knee is an example of a sesamoid bone.

Sutural bones are small, flat, irregularly shaped bones. They may be found between the flat bones of the skull. They vary in number, shape, size, and position.

Bone Tissue

Bones are considered organs because they contain various types of tissue, such as blood, connective tissue, nerves, and bone tissue. Osteocytes, the living cells of bone tissue, form the mineral matrix of bones. There are two types of bone tissue: compact and spongy.

Compact Bone Tissue

Compact bone (or cortical bone) forms the hard external layer of all bones and surrounds the medullary cavity, or bone marrow. It provides protection and strength to bones. Compact bone tissue consists of units called osteons or Haversian systems. **Osteons** are cylindrical structures that contain a mineral matrix and living osteocytes connected by canaliculi, which transport blood. They are aligned parallel to the long axis of the bone. Each osteon consists of **lamellae**, which are layers of compact matrix that surround a central canal called the Haversian canal. The **Haversian canal (osteonic canal)** contains the bone's blood vessels and nerve fibers (Figure 19.19). Osteons in compact bone tissue are aligned in the same direction along lines of stress and help the bone resist bending or fracturing. Therefore, compact bone tissue is prominent in areas of bone at which stresses are applied in only a few directions.

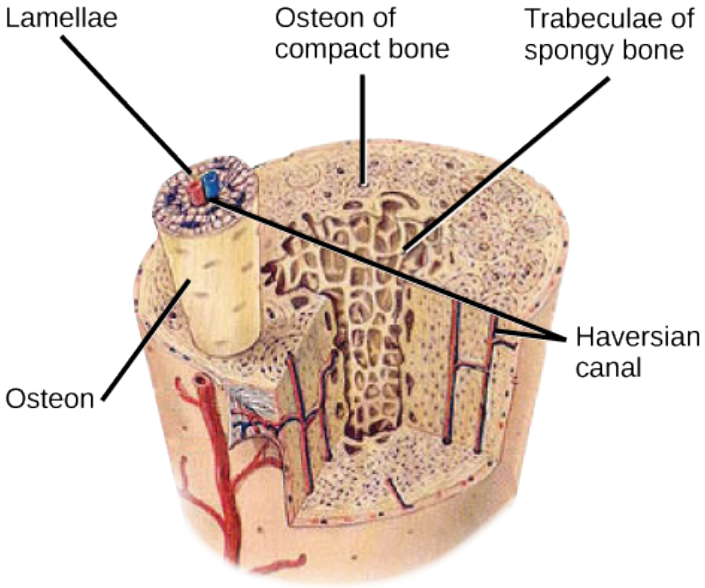


Figure 19.19. Compact bone tissue consists of osteons that are aligned parallel to the long axis of the bone, and the Haversian canal that contains the bone's blood vessels and nerve fibers. The inner layer of bones consists of spongy bone tissue. The small dark ovals in the osteon represent the living osteocytes. (credit: modification of work by NCI, NIH)

Which of the following statements about bone tissue is false?

1. Compact bone tissue is made of cylindrical osteons that are aligned such that they travel the length of the bone.
2. Haversian canals contain blood vessels only.
3. Haversian canals contain blood vessels and nerve fibers.
4. Spongy tissue is found on the interior of the bone, and compact bone tissue is found on the exterior.

Spongy Bone Tissue

Whereas compact bone tissue forms the outer layer of all bones, **spongy bone** or cancellous bone forms the inner layer of all bones. Spongy bone tissue does not contain osteons that constitute compact bone tissue. Instead, it consists of **trabeculae**, which are lamellae that are arranged as rods or plates. Red bone marrow is found between the trabuculae. Blood vessels within this tissue deliver nutrients to osteocytes and remove waste. The red bone marrow of the femur and the interior of other large bones, such as the ileum, forms blood cells.

Spongy bone reduces the density of bone and allows the ends of long bones to compress as the result of stresses applied to the bone. Spongy bone is prominent in areas of bones that are not heavily stressed or where stresses arrive from many directions. The epiphyses of bones, such as the neck of the femur, are subject to stress from many directions. Imagine laying a heavy framed picture flat on the floor. You

could hold up one side of the picture with a toothpick if the toothpick was perpendicular to the floor and the picture. Now drill a hole and stick the toothpick into the wall to hang up the picture. In this case, the function of the toothpick is to transmit the downward pressure of the picture to the wall. The force on the picture is straight down to the floor, but the force on the toothpick is both the picture wire pulling down and the bottom of the hole in the wall pushing up. The toothpick will break off right at the wall.

The neck of the femur is horizontal like the toothpick in the wall. The weight of the body pushes it down near the joint, but the vertical diaphysis of the femur pushes it up at the other end. The neck of the femur must be strong enough to transfer the downward force of the body weight horizontally to the vertical shaft of the femur (Figure 19.20).

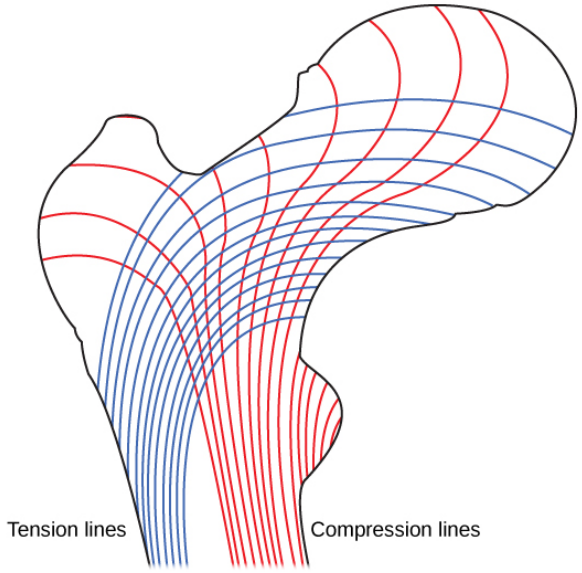


Figure 19.20.
Trabeculae in spongy bone are arranged such that one side of the bone bears tension and the other withstands compression.

Concept in Action

View micrographs of musculoskeletal tissues as you review the anatomy.

Cell Types in Bones

Bone consists of four types of cells: osteoblasts, osteoclasts, osteocytes, and osteoprogenitor cells.

Osteoblasts are bone cells that are responsible for bone formation. Osteoblasts synthesize and secrete the organic part and inorganic part of the extracellular matrix of bone tissue, and collagen fibers. Osteoblasts become trapped in these secretions and differentiate into less active osteocytes. **Osteoclasts** are large bone cells with up to 50 nuclei. They remove bone structure by releasing lysosomal enzymes and acids that dissolve the bony matrix. These minerals, released from bones into the blood, help regulate calcium concentrations in body fluids. Bone may also be resorbed for remodeling, if the applied stresses have changed. **Osteocytes** are mature bone cells and are the main cells in bony connective tissue; these cells cannot divide. Osteocytes maintain normal bone structure by recycling the mineral salts in the bony matrix. **Osteoprogenitor cells** are squamous stem cells that divide to produce daughter cells that differentiate into osteoblasts. Osteoprogenitor cells are important in the repair of fractures.

Development of Bone

Ossification, or osteogenesis, is the process of bone formation by osteoblasts. Ossification is distinct from the process of calcification; whereas calcification takes place during the ossification of bones, it can also occur in other tissues. Ossification begins approximately six weeks after fertilization in an embryo. Before this time, the embryonic skeleton consists entirely of fibrous membranes and hyaline cartilage. The development of bone from fibrous membranes is called intramembranous ossification; development from hyaline cartilage is called endochondral ossification. Bone growth continues until approximately age 25. Bones can grow in thickness throughout life, but after age 25, ossification functions primarily in bone remodeling and repair.

Intramembranous Ossification

Intramembranous ossification is the process of bone development from fibrous membranes. It is involved in the formation of the flat bones of the skull, the mandible, and the clavicles. Ossification begins as mesenchymal cells form a template of the future bone. They then differentiate into osteoblasts at the ossification center. Osteoblasts secrete the extracellular matrix and deposit calcium, which hardens the matrix. The non-mineralized portion of the bone or osteoid continues to form around blood vessels, forming spongy bone. Connective tissue in the matrix differentiates into red bone marrow in the fetus. The spongy bone is remodeled into a thin layer of compact bone on the surface of the spongy bone.

Endochondral Ossification

Endochondral ossification

is the process of bone development from hyaline cartilage. All of the bones of the body, except for the flat bones of the skull, mandible, and clavicles, are formed through endochondral ossification.

In long bones, chondrocytes form a template of the hyaline cartilage diaphysis. Responding to complex developmental signals, the matrix begins to calcify. This calcification prevents diffusion of nutrients into the matrix, resulting in chondrocytes dying and the opening up of cavities in the diaphysis cartilage. Blood vessels invade the cavities, and osteoblasts and osteoclasts modify the calcified

cartilage matrix into spongy bone. Osteoclasts then break down some of the spongy bone to create a marrow, or medullary, cavity in the center of the diaphysis. Dense, irregular connective tissue forms a sheath (periosteum) around the bones. The periosteum assists in attaching the bone to surrounding tissues, tendons, and ligaments. The bone continues to grow and elongate as the cartilage cells at the epiphyses divide.

In the last stage of prenatal bone development, the centers of the epiphyses begin to calcify. Secondary ossification centers form in the epiphyses as blood vessels and osteoblasts enter these areas and convert hyaline cartilage into spongy bone. Until adolescence, hyaline cartilage persists at the **epiphyseal plate** (growth plate), which is the region between the diaphysis and epiphysis that is responsible for the lengthwise growth of long bones (Figure 19.21).

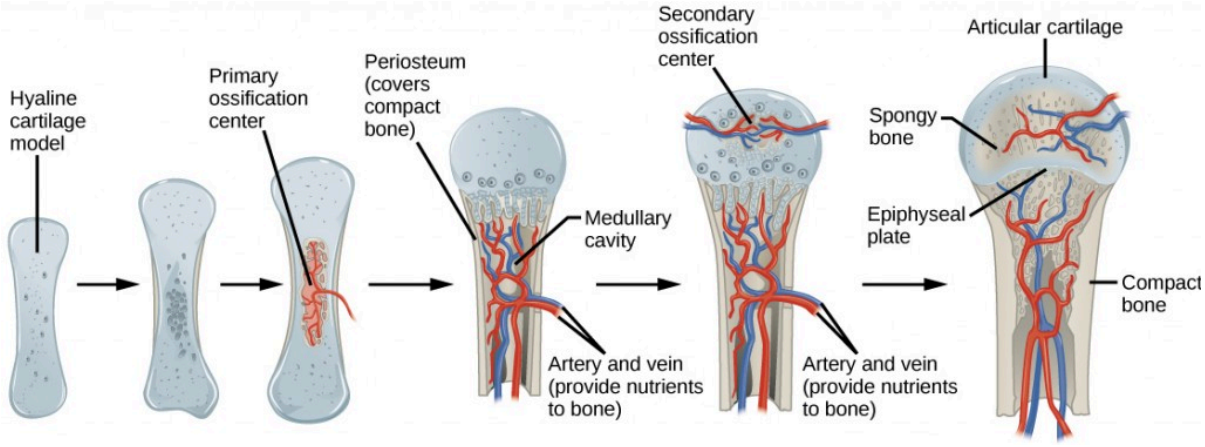


Figure 19.21.
Endochondral ossification is the process of bone development from hyaline cartilage. The periosteum is the connective tissue on the outside of bone that acts as the interface between bone, blood vessels, tendons, and ligaments.

Growth of Bone

Long bones continue to lengthen, potentially until adolescence, through the addition of bone tissue at the epiphyseal plate. They also increase in width through appositional growth.

Lengthening of Long Bones

Chondrocytes on the epiphyseal side of the epiphyseal plate divide; one cell remains undifferentiated near the epiphysis, and one cell moves toward the diaphysis. The cells, which are pushed from the epiphysis, mature and are destroyed by calcification. This process replaces cartilage with bone on the diaphyseal side of the plate, resulting in a lengthening of the bone.

Long bones stop growing at around the age of 18 in females and the age of 21 in males in a process called epiphyseal plate closure. During this process, cartilage cells stop dividing and all of the cartilage is replaced by bone. The epiphyseal plate fades, leaving a structure called the epiphyseal line or epiphyseal remnant, and the epiphysis and diaphysis fuse.

Thickening of Long Bones

Appositional growth is the increase in the diameter of bones by the addition of bony tissue at the surface of bones. Osteoblasts at the bone surface secrete bone matrix, and osteoclasts on the inner surface break down bone. The osteoblasts differentiate into osteocytes. A balance between these two processes allows the bone to thicken without becoming too heavy.

Bone Remodeling and Repair

Bone renewal continues after birth into adulthood. **Bone remodeling** is the replacement of old bone tissue by new bone tissue. It involves the processes of bone deposition by osteoblasts and bone resorption by osteoclasts. Normal bone growth requires vitamins D, C, and A, plus minerals such as calcium, phosphorous, and magnesium. Hormones such as parathyroid hormone, growth hormone, and calcitonin are also required for proper bone growth and maintenance.

Bone turnover rates are quite high, with five to seven percent of bone mass being recycled every week. Differences in turnover rate exist in different areas of the skeleton and in different areas of a bone. For example, the bone in the head of the femur may be fully replaced every six months, whereas the bone along the shaft is altered much more slowly.

Bone remodeling allows bones to adapt to stresses by becoming thicker and stronger when subjected to stress. Bones that are not subject to normal stress, for example when a limb is in a cast, will begin to lose mass. A fractured or broken bone undergoes repair through four stages:

1. Blood vessels in the broken bone tear and hemorrhage, resulting in the formation of clotted blood, or a hematoma, at the site of the break. The severed blood vessels at the broken ends of the bone are sealed by the clotting process, and bone cells that are deprived of nutrients begin to die.
2. Within days of the fracture, capillaries grow into the hematoma, and phagocytic cells begin to clear away the dead cells. Though fragments of the blood clot may remain, fibroblasts and osteoblasts enter the area and begin to reform bone. Fibroblasts produce collagen fibers that connect the broken bone ends, and osteoblasts start to form spongy bone. The repair tissue between the broken bone ends is called the fibrocartilaginous callus, as it is composed of both hyaline and fibrocartilage (Figure 19.22). Some bone spicules may also appear at this point.
3. The fibrocartilaginous callus is converted into a bony callus of spongy bone. It takes about two months for the broken bone ends to be firmly joined together after the fracture. This is similar to the endochondral formation of bone, as cartilage becomes ossified; osteoblasts, osteoclasts, and bone matrix are present.
4. The bony callus is then remodelled by osteoclasts and osteoblasts, with excess material on the exterior of the bone and within the medullary cavity being removed. Compact bone is added to create bone tissue that is similar to the original, unbroken bone. This remodeling can take many months, and the bone may remain uneven for years.

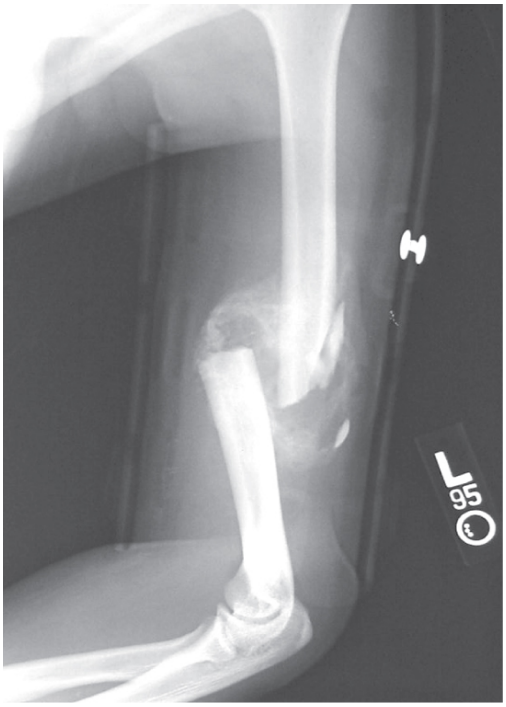


Figure 19.22. After this bone is set, a callus will knit the two ends together. (credit: Bill Rhodes)

Scientific Method Connection

Decalcification of Bones

Question: What effect does the removal of calcium and collagen have on bone structure?

Background: Conduct a literature search on the role of calcium and collagen in maintaining bone structure. Conduct a literature search on diseases in which bone structure is compromised.

Hypothesis: Develop a hypothesis that states predictions of the flexibility, strength, and mass of bones that have had the calcium and collagen components removed. Develop a hypothesis regarding the attempt to add calcium back to decalcified bones.

Test the hypothesis: Test the prediction by removing calcium from chicken bones by placing them in a jar of vinegar for seven days. Test the hypothesis regarding adding calcium back to decalcified bone by placing the decalcified chicken bones into a jar of water with calcium supplements added. Test the prediction by denaturing the collagen from the bones by baking them at 250°C for three hours.

Analyze the data: Create a table showing the changes in bone flexibility, strength, and mass in the three different environments.

Report the results: Under which conditions was the bone most flexible? Under which conditions was the bone the strongest?

Draw a conclusion: Did the results support or refute the hypothesis? How do the results observed in this experiment correspond to diseases that destroy bone tissue?

Summary

Bone, or osseous tissue, is connective tissue that includes specialized cells, mineral salts, and collagen fibers. The human skeleton can be divided into long bones, short bones, flat bones, and irregular bones. Compact bone tissue is composed of osteons and forms the external layer of all bones. Spongy bone tissue is composed of trabeculae and forms the inner part of all bones. Four types of cells compose bony tissue: osteocytes, osteoclasts, osteoprogenitor cells, and osteoblasts. Ossification is the process of bone formation by osteoblasts. Intramembranous ossification is the process of bone development from fibrous membranes. Endochondral ossification is the process of bone development from hyaline cartilage. Long bones lengthen as chondrocytes divide and secrete hyaline cartilage. Osteoblasts replace cartilage with bone. Appositional growth is the increase in the diameter of bones by the addition of bone tissue at the surface of bones. Bone remodeling involves the processes of bone deposition by osteoblasts and bone resorption by osteoclasts. Bone repair occurs in four stages and can take several months.

Exercises

1. The Haversian canal:
    1. is arranged as rods or plates
    2. contains the bone’s blood vessels and nerve fibers
    3. is responsible for the lengthwise growth of long bones
    4. synthesizes and secretes matrix
2. The epiphyseal plate:
    1. is arranged as rods or plates
    2. contains the bone’s blood vessels and nerve fibers
    3. is responsible for the lengthwise growth of long bones
    4. synthesizes and secretes bone matrix
3. The cells responsible for bone resorption are _____.
    1. osteoclasts
    2. osteoblasts
    3. fibroblasts
    4. osteocytes
4. Compact bone is composed of _____.
    1. trabeculae

2. compacted collagen
3. osteons
4. calcium phosphate only

5. What are the major differences between spongy bone and compact bone?
6. What are the roles of osteoblasts, osteocytes, and osteoclasts?

Answers

1. B
2. C
3. A
4. C
5. Compact bone tissue forms the hard external layer of all bones and consists of osteons. Compact bone tissue is prominent in areas of bone at which stresses are applied in only a few directions. Spongy bone tissue forms the inner layer of all bones and consists of trabeculae. Spongy bone is prominent in areas of bones that are not heavily stressed or at which stresses arrive from many directions.
6. Osteocytes function in the exchange of nutrients and wastes with the blood. They also maintain normal bone structure by recycling the mineral salts in the bony matrix. Osteoclasts remove bone tissue by releasing lysosomal enzymes and acids that dissolve the bony matrix. Osteoblasts are bone cells that are responsible for bone formation.

Glossary

appositional growth
increase in the diameter of bones by the addition of bone tissue at the surface of bones

bone remodeling
replacement of old bone tissue by new bone tissue

bone
(also, osseous tissue) connective tissue that constitutes the endoskeleton

calcification
process of deposition of mineral salts in the collagen fiber matrix that crystallizes and hardens the tissue

compact bone
forms the hard external layer of all bones

diaphysis
central shaft of bone, contains bone marrow in a marrow cavity

endochondral ossification
process of bone development from hyaline cartilage

epiphyseal plate
region between the diaphysis and epiphysis that is responsible for the lengthwise growth of long bones

epiphysis
rounded end of bone, covered with articular cartilage and filled with red bone marrow, which produces blood cells

flat bone
thin and relatively broad bone found where extensive protection of organs is required or where broad surfaces of muscle attachment are required

Haversian canal
contains the bone's blood vessels and nerve fibers

intramembranous ossification
process of bone development from fibrous membranes

irregular bone
bone with complex shapes; examples include vertebrae and hip bones

lamella
layer of compact tissue that surrounds a central canal called the Haversian canal

long bone
bone that is longer than wide, and has a shaft and two ends

osseous tissue
connective tissue that constitutes the endoskeleton

ossification
(also, osteogenesis) process of bone formation by osteoblasts

osteoblast
bone cell responsible for bone formation

osteoclast
large bone cells with up to 50 nuclei, responsible for bone remodeling

osteocyte
mature bone cells and the main cell in bone tissue

osteon
cylindrical structure aligned parallel to the long axis of the bone

sesamoid bone
small, flat bone shaped like a sesame seed; develops inside tendons

short bone
bone that has the same width and length, giving it a cube-like shape

spongy bone tissue
forms the inner layer of all bones

suture bone
small, flat, irregularly shaped bone that forms between the flat bones of the cranium

trabeculae
lamellae that are arranged as rods or plates

19.3 Joints and Skeletal Movement

Learning Objectives

By the end of this section, you will be able to:

- Classify the different types of joints on the basis of structure
- Explain the role of joints in skeletal movement

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

Classification of Joints on the Basis of Structure

here are two ways to classify joints: on the basis of their structure or on the basis of their function. The structural classification divides joints into bony, fibrous, cartilaginous, and synovial joints depending on the material composing the joint and the presence or absence of a cavity in the joint.

Fibrous Joints

The bones of **fibrous joints** are held together by fibrous connective tissue. There is no cavity, or space, present between the bones and so most fibrous joints do not move at all, or are only capable of minor movements. There are three types of fibrous joints: sutures, syndesmoses, and gomphoses. **Sutures** are found only in the skull and possess short fibers of connective tissue that hold the skull bones tightly in place (Figure 19.23).

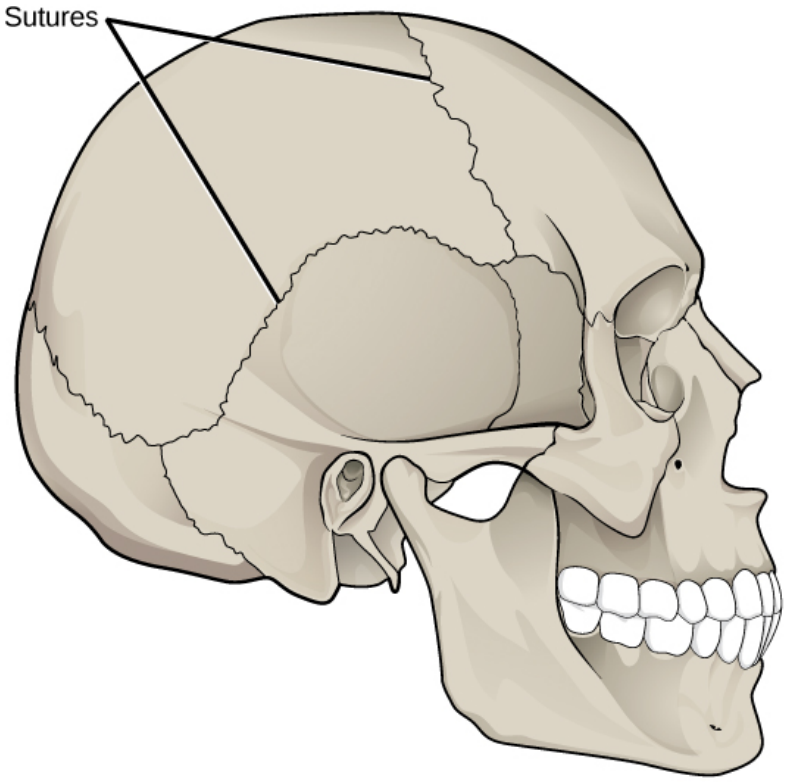


Figure 19.23. Sutures are fibrous joints found only in the skull.

Syndesmoses are joints in which the bones are connected by a band of connective tissue, allowing for more movement than in a suture. An example of a syndesmosis is the joint of the tibia and fibula in the ankle. The amount of movement in these types of joints is determined by the length of the connective tissue fibers. **Gomphoses** occur between teeth and their sockets; the term refers to the way the tooth fits into the socket like a peg (Figure 19.24). The tooth is connected to the socket by a connective tissue referred to as the periodontal ligament.

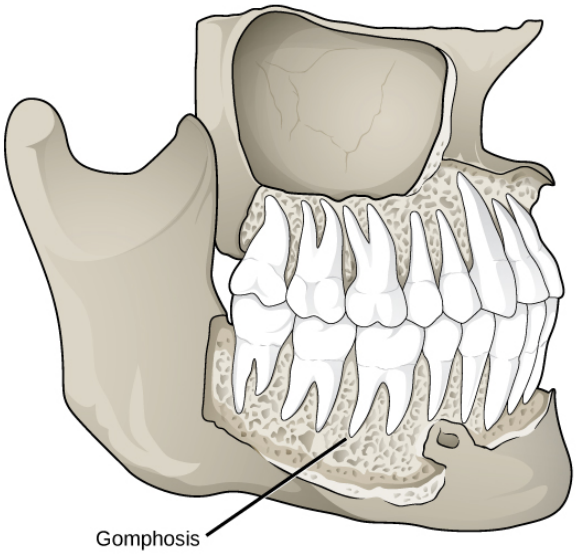


Figure 19.24. Gomphoses are fibrous joints between the teeth and their sockets. (credit: modification of work by Gray's Anatomy)

Cartilaginous Joints

Cartilaginous joints are joints in which the bones are connected by cartilage. There are two types of cartilaginous joints: synchondroses and symphyses. In a **synchondrosis**, the bones are joined by hyaline cartilage. Synchondroses are found in the epiphyseal plates of growing bones in children. In **symphyses**, hyaline cartilage covers the end of the bone but the connection between bones occurs through fibrocartilage. Symphyses are found at the joints between vertebrae. Either type of cartilaginous joint allows for very little movement.

Synovial Joints

Synovial joints are the only joints that have a space between the adjoining bones (Figure 19.25). This space is referred to as the synovial (or joint) cavity and is filled with synovial fluid. Synovial fluid lubricates the joint, reducing friction between the bones and allowing for greater movement. The ends of the bones are covered with articular cartilage, a hyaline cartilage, and the entire joint is surrounded by an articular capsule composed of connective tissue that allows movement of the joint while resisting dislocation. Articular capsules may also possess ligaments that hold the bones together. Synovial joints are capable of the greatest movement of the three structural joint types; however, the more mobile a joint, the weaker the joint. Knees, elbows, and shoulders are examples of synovial joints.

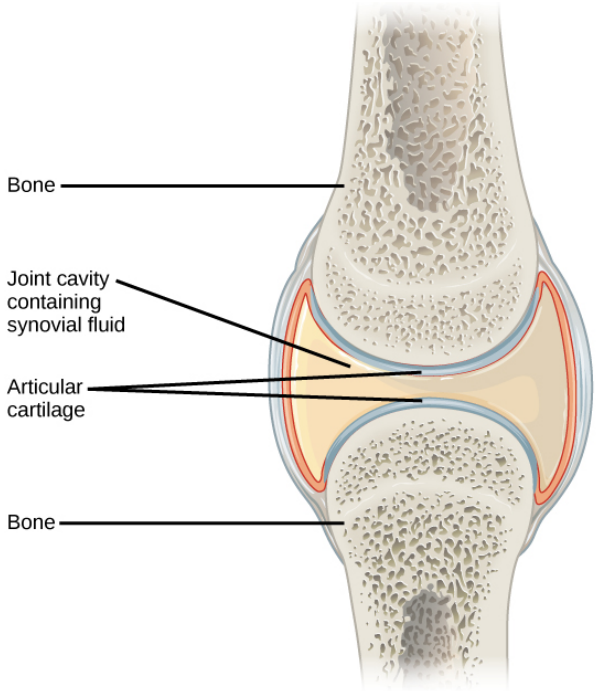


Figure 19.25.
Synovial joints are the only joints that have a space or "synovial cavity" in the joint.

Classification of Joints on the Basis of Function

The functional classification divides joints into three categories: synarthroses, amphiarthroses, and diarthroses. A **synarthrosis** is a joint that is immovable. This includes sutures, gomphoses, and

synchondroses. **Amphiarthroses** are joints that allow slight movement, including syndesmoses and symphyses. **Diarthroses**are joints that allow for free movement of the joint, as in synovial joints.

Movement at Synovial Joints

The wide range of movement allowed by synovial joints produces different types of movements. The movement of synovial joints can be classified as one of four different types: gliding, angular, rotational, or special movement.

Gliding Movement

Gliding movementsoccur as relatively flat bone surfaces move past each other. Gliding movements produce very little rotation or angular movement of the bones. The joints of the carpal and tarsal bones are examples of joints that produce gliding movements.

Angular Movement

Angular movements are produced when the angle between the bones of a joint changes. There are several different types of angular movements, including flexion, extension, hyperextension, abduction, adduction, and circumduction. **Flexion**, or bending, occurs when the angle between the bones decreases. Moving the forearm upward at the elbow or moving the wrist to move the hand toward the forearm are examples of flexion.**Extension** is the opposite of flexion in that the angle between the bones of a joint increases. Straightening a limb after flexion is an example of extension. Extension past the regular anatomical position is referred to as **hyperextension**. This includes moving the neck back to look upward, or bending the wrist so that the hand moves away from the forearm.

Abduction occurs when a bone moves away from the midline of the body. Examples of abduction are moving the arms or legs laterally to lift them straight out to the side. **Adduction**is the movement of a bone toward the midline of the body. Movement of the limbs inward after abduction is an example of adduction.**Circumduction** is the movement of a limb in a circular motion, as in moving the arm in a circular motion.

Rotational Movement

Rotational movement is the movement of a bone as it rotates around its longitudinal axis. Rotation can be toward the midline of the body, which is referred to as **medial rotation**, or away from the midline of the body, which is referred to as **lateral rotation**. Movement of the head from side to side is an example of rotation.

Special Movements

Some movements that cannot be classified as gliding, angular, or rotational are called special movements. **Inversion** involves the soles of the feet moving inward, toward the midline of the body. **Eversion** is the opposite of inversion, movement of the sole of the foot outward, away from the midline of the body. **Protraction** is the anterior movement of a bone in the horizontal plane. **Retraction**occurs

as a joint moves back into position after protraction. Protraction and retraction can be seen in the movement of the mandible as the jaw is thrust outwards and then back inwards. **Elevation** is the movement of a bone upward, such as when the shoulders are shrugged, lifting the scapulae. **Depression** is the opposite of elevation—movement downward of a bone, such as after the shoulders are shrugged and the scapulae return to their normal position from an elevated position.**Dorsiflexion** is a bending at the ankle such that the toes are lifted toward the knee. **Plantar flexion** is a bending at the ankle when the heel is lifted, such as when standing on the toes. **Supination** is the movement of the radius and ulna bones of the forearm so that the palm faces forward. **Pronation** is the opposite movement, in which the palm faces backward. **Opposition** is the movement of the thumb toward the fingers of the same hand, making it possible to grasp and hold objects.

Types of Synovial Joints

Synovial joints are further classified into six different categories on the basis of the shape and structure of the joint. The shape of the joint affects the type of movement permitted by the joint (Figure 19.26). These joints can be described as planar, hinge, pivot, condyloid, saddle, or ball-and-socket joints.

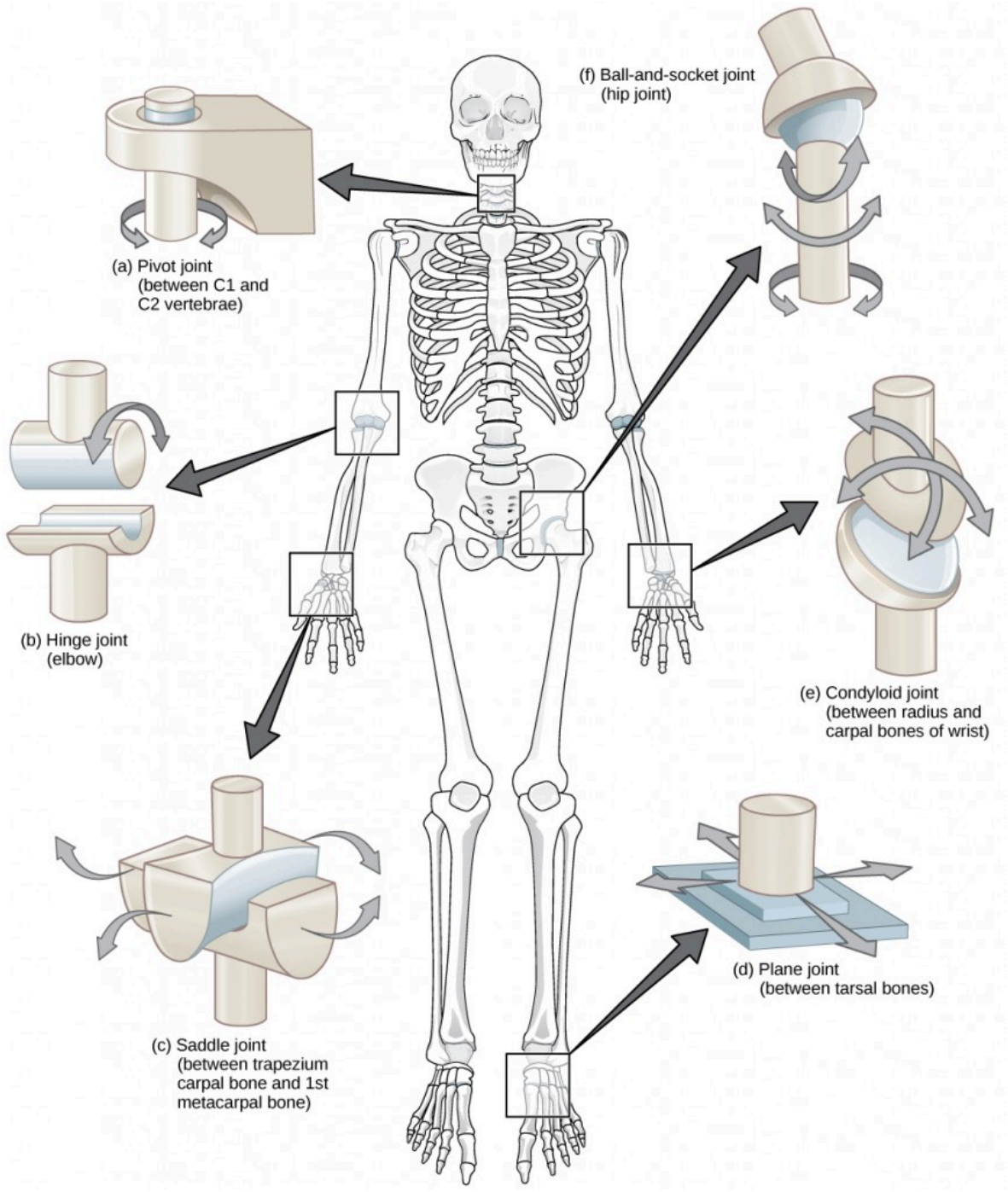


Figure 19.26.
Different types of joints allow different types of movement. Planar, hinge, pivot, condyloid, saddle, and ball-and-socket are all types of synovial joints.

Planar Joints

Planar joints have bones with articulating surfaces that are flat or slightly curved faces. These joints allow for gliding movements, and so the joints are sometimes referred to as gliding joints. The range of motion is limited in these joints and does not involve rotation. Planar joints are found in the carpal bones in the hand and the tarsal bones of the foot, as well as between vertebrae (Figure 19.27).

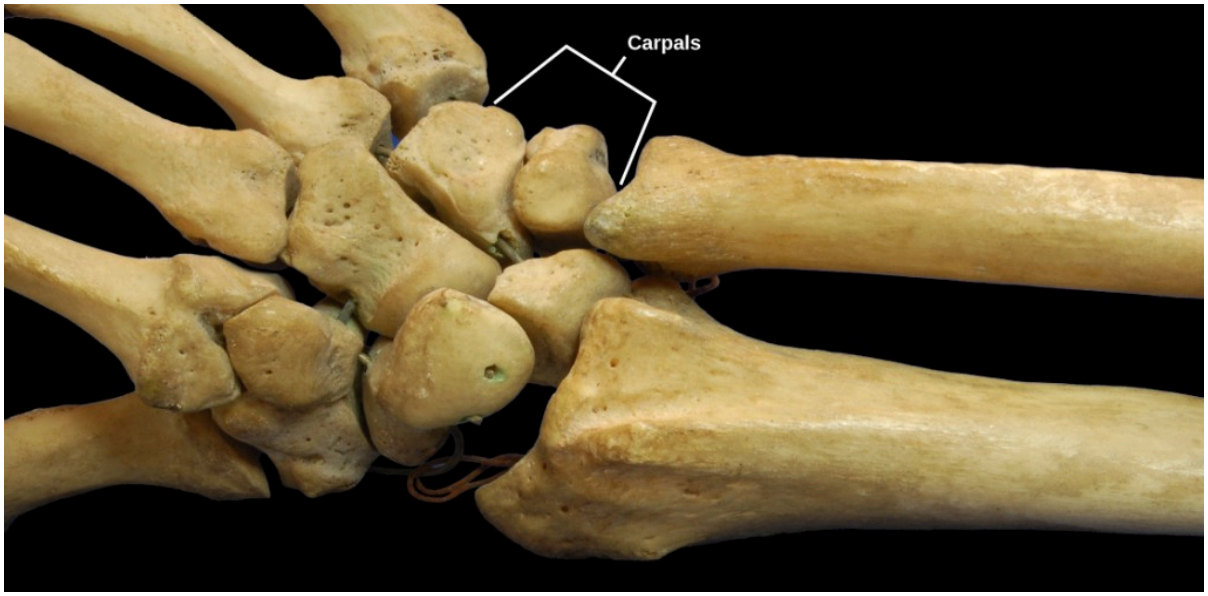


Figure 19.27.
The joints of the carpal bones in the wrist are examples of planar joints. (credit: modification of work by Brian C. Goss)

Hinge Joints

In **hinge joints**, the slightly rounded end of one bone fits into the slightly hollow end of the other bone. In this way, one bone moves while the other remains stationary, like the hinge of a door. The elbow is an example of a hinge joint. The knee is sometimes classified as a modified hinge joint (Figure 19.28).

Figure 19.28.
The elbow joint, where the radius articulates with the humerus, is an example of a hinge joint. (credit: modification of work by Brian C. Goss)

Pivot Joints

Pivot joints consist of the rounded end of one bone fitting into a ring formed by the other bone. This structure allows rotational movement, as the rounded bone moves around its own axis. An example of a pivot joint is the joint of the first and second vertebrae of the neck that allows the head to move back

and forth (Figure 19.29). The joint of the wrist that allows the palm of the hand to be turned up and down is also a pivot joint.

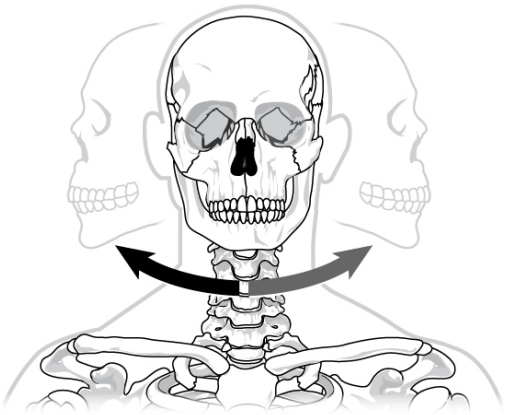


Figure 19.29. The joint in the neck that allows the head to move back and forth is an example of a pivot joint.

Condyloid Joints

Condyloid joints consist of an oval-shaped end of one bone fitting into a similarly oval-shaped hollow of another bone (Figure 19.30). This is also sometimes called an ellipsoidal joint. This type of joint allows angular movement along two axes, as seen in the joints of the wrist and fingers, which can move both side to side and up and down.

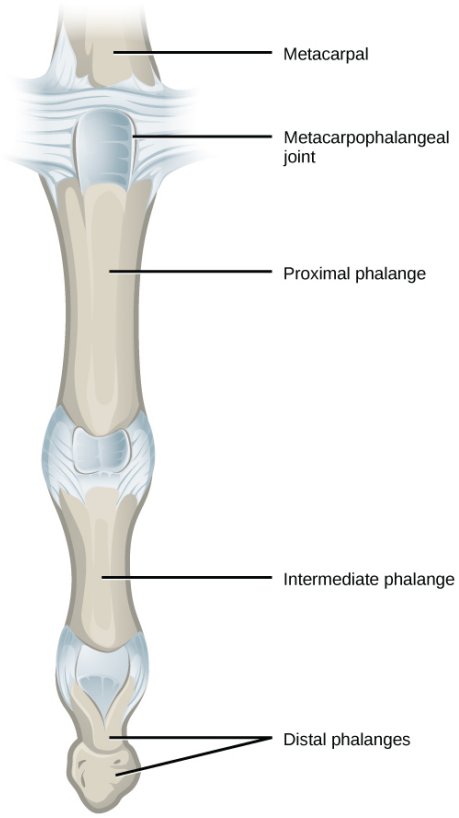


Figure 19.30.
The metacarpophalangeal joints in the finger are examples of condyloid joints. (credit: modification of work by Gray's Anatomy)

Saddle Joints

Saddle joints are so named because the ends of each bone resemble a saddle, with concave and convex portions that fit together. Saddle joints allow angular movements similar to condyloid joints but with a greater range of motion. An example of a saddle joint is the thumb joint, which can move back and forth and up and down, but more freely than the wrist or fingers (Figure 19.31).

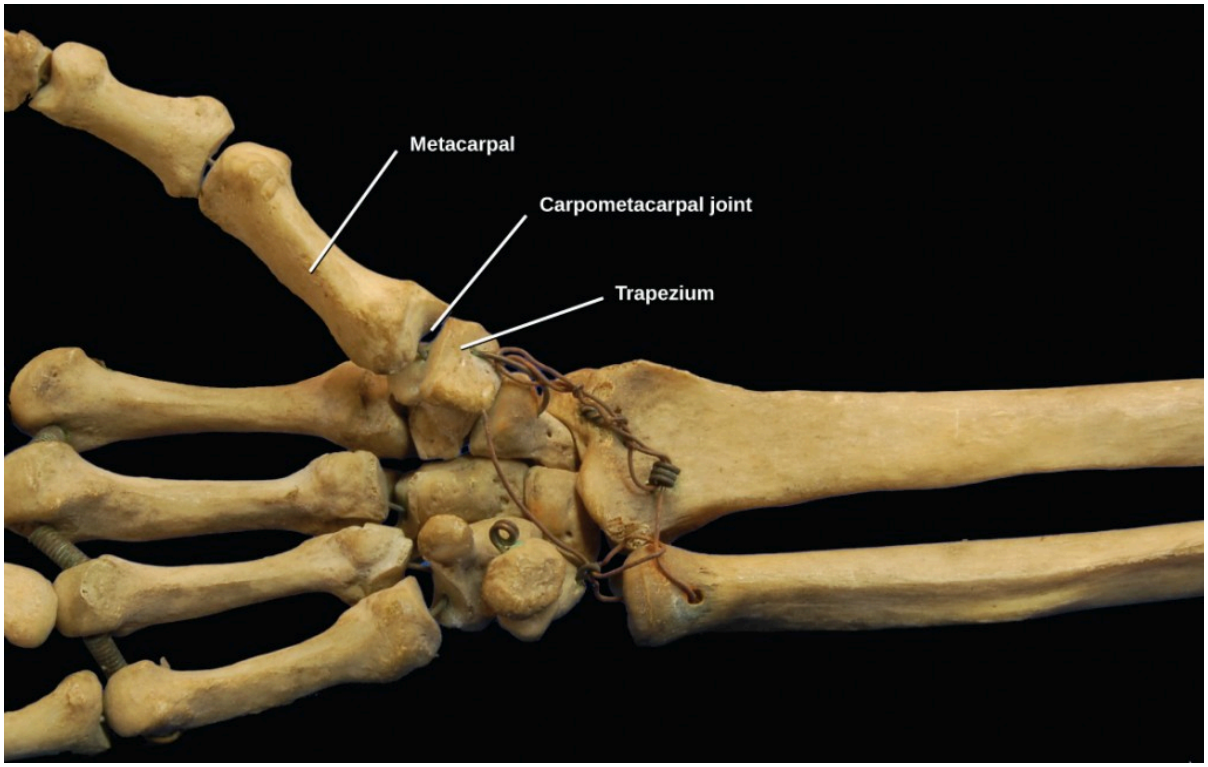


Figure 19.31.
The carpometacarpal joints in the thumb are examples of saddle joints. (credit: modification of work by Brian C. Goss)

Ball-and-Socket Joints

Ball-and-socket joints possess a rounded, ball-like end of one bone fitting into a cuplike socket of another bone. This organization allows the greatest range of motion, as all movement types are possible in all directions. Examples of ball-and-socket joints are the shoulder and hip joints (Figure 19.32).

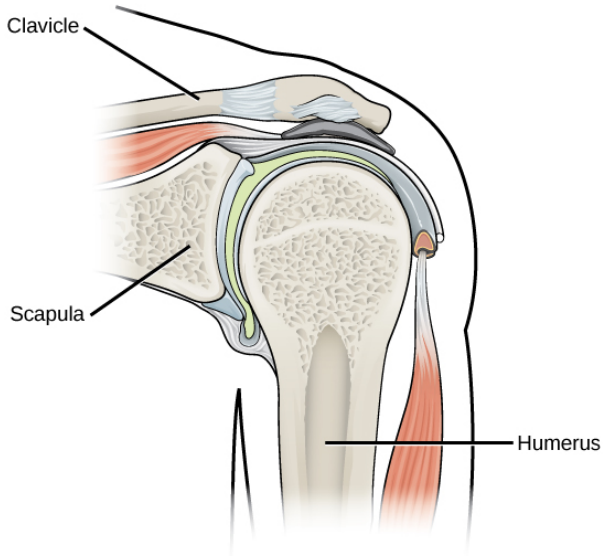


Figure 19.32.
The shoulder joint is an example of a ball-and-socket joint.

Concept in Action

Watch this animation showing the six types of synovial joints.

Rheumatologist

Rheumatologists are medical doctors who specialize in the diagnosis and treatment of disorders of the joints, muscles, and bones. They diagnose and treat diseases such as arthritis, musculoskeletal disorders, osteoporosis, and autoimmune diseases such as ankylosing spondylitis and rheumatoid arthritis.

Rheumatoid arthritis (RA) is an inflammatory disorder that primarily affects the synovial joints of the hands, feet, and cervical spine. Affected joints become swollen, stiff, and painful. Although it is known that RA is an autoimmune disease in which the body's immune system mistakenly attacks healthy tissue, the cause of RA remains unknown. Immune cells from the blood enter joints and the synovium causing cartilage breakdown, swelling, and inflammation of the joint lining. Breakdown of cartilage causes bones to rub against each other causing pain. RA is more common in women than men and the age of onset is usually 40–50 years of age.

Rheumatologists can diagnose RA on the basis of symptoms such as joint inflammation and pain, X-

ray and MRI imaging, and blood tests. Arthrography is a type of medical imaging of joints that uses a contrast agent, such as a dye, that is opaque to X-rays. This allows the soft tissue structures of joints—such as cartilage, tendons, and ligaments—to be visualized. An arthrogram differs from a regular X-ray by showing the surface of soft tissues lining the joint in addition to joint bones. An arthrogram allows early degenerative changes in joint cartilage to be detected before bones become affected.

There is currently no cure for RA; however, rheumatologists have a number of treatment options available. Early stages can be treated with rest of the affected joints by using a cane or by using joint splints that minimize inflammation. When inflammation has decreased, exercise can be used to strengthen the muscles that surround the joint and to maintain joint flexibility. If joint damage is more extensive, medications can be used to relieve pain and decrease inflammation. Anti-inflammatory drugs such as aspirin, topical pain relievers, and corticosteroid injections may be used. Surgery may be required in cases in which joint damage is severe.

Summary

The structural classification of joints divides them into bony, fibrous, cartilaginous, and synovial joints. The bones of fibrous joints are held together by fibrous connective tissue; the three types of fibrous joints are sutures, syndesomes, and gomphoses. Cartilaginous joints are joints in which the bones are connected by cartilage; the two types of cartilaginous joints are synchondroses and symphyses. Synovial joints are joints that have a space between the adjoining bones. The functional classification divides joints into three categories: synarthroses, amphiarthroses, and diarthroses. The movement of synovial joints can be classified as one of four different types: gliding, angular, rotational, or special movement. Gliding movements occur as relatively flat bone surfaces move past each other. Angular movements are produced when the angle between the bones of a joint changes. Rotational movement is the movement of a bone as it rotates around its own longitudinal axis. Special movements include inversion, eversion, protraction, retraction, elevation, depression, dorsiflexion, plantar flexion, supination, pronation, and opposition. Synovial joints are also classified into six different categories on the basis of the shape and structure of the joint: planar, hinge, pivot, condyloid, saddle, and ball-and-socket.

Exercises

1. Synchondroses and symphyses are:
    1. synovial joints
    2. cartilaginous joints
    3. fibrous joints
    4. condyloid joints
2. The movement of bone away from the midline of the body is called _____.
    1. circumduction

2. extension
3. adduction
4. abductino

3. Which of the following is not a characteristic of the synovial fluid?

1. lubrication
2. shock absorption
3. regulation of water balance in the joint
4. protection of articular cartilage

4. The elbow is an example of which type of joint?

1. hinge
2. pivot
3. saddle
4. gliding

5. What movements occur at the hip joint and knees as you bend down to touch your toes?
6. What movement(s) occur(s) at the scapulae when you shrug your shoulders?

Answers

1. B
2. D
3. C
4. A
5. The hip joint is flexed and the knees are extended.
6. Elevation is the movement of a bone upward, such as when the shoulders are shrugged, lifting the scapulae. Depression is the downward movement of a bone, such as after the shoulders are shrugged and the scapulae return to their normal position from an elevated position.

Glossary

abduction
when a bone moves away from the midline of the body

adduction
movement of the limbs inward after abduction

amphiarthrosis
joint that allows slight movement; includes syndesmoses and symphyses

angular movement
produced when the angle between the bones of a joint changes

articulation
any place where two bones are joined

ball-and-socket joint
joint with a rounded, ball-like end of one bone fitting into a cuplike socket of another bone

cartilaginous joint
joint in which the bones are connected by cartilage

circumduction
movement of a limb in a circular motion.

condyloid joint
oval-shaped end of one bone fitting into a similarly oval-shaped hollow of another bone

depression
movement downward of a bone, such as after the shoulders are shrugged and the scapulae return to their normal position from an elevated position; opposite of elevation

diarthrosis
joint that allows for free movement of the joint; found in synovial joints

dorsiflexion
bending at the ankle such that the toes are lifted toward the knee

elevation
movement of a bone upward, such as when the shoulders are shrugged, lifting the scapulae

eversion
movement of the sole of the foot outward, away from the midline of the body; opposite of inversion

extension
movement in which the angle between the bones of a joint increases; opposite of flexion

fibrous joint
joint held together by fibrous connective tissue

flexion
movement in which the angle between the bones decreases; opposite of extension

gliding movement
when relatively flat bone surfaces move past each other

gomphosis
the joint in which the tooth fits into the socket like a peg

hinge joint
slightly rounded end of one bone fits into the slightly hollow end of the other bone

inversion
soles of the feet moving inward, toward the midline of the body

joint
point at which two or more bones meet

lateral rotation
rotation away from the midline of the body

medial rotation
rotation toward the midline of the body

opposition
movement of the thumb toward the fingers of the same hand, making it possible to grasp and hold objects

pivot joint
joint with the rounded end of one bone fitting into a ring formed by the other bone

planar joint
joint with bones whose articulating surfaces are flat

plantar flexion
bending at the ankle such that the heel is lifted, such as when standing on the toes

pronation
movement in which the palm faces backward

protraction
anterior movement of a bone in the horizontal plane

retraction
movement in which a joint moves back into position after protraction

rotational movement
movement of a bone as it rotates around its own longitudinal axis

saddle joint
joint with concave and convex portions that fit together; named because the ends of each bone resemble a saddle

supination
movement of the radius and ulna bones of the forearm so that the palm faces forward

synarthrosis
joint that is immovable

syndesmosis
joint in which the bones are connected by a band of connective tissue, allowing for more movement than in a suture

synovial joint
only joint that has a space between the adjoining bones

19.4 Muscle Contraction and Locomotion

Learning Objectives

By the end of this section, you will be able to:

- Classify the different types of muscle tissue
- Explain the role of muscles in locomotion

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 19.33).

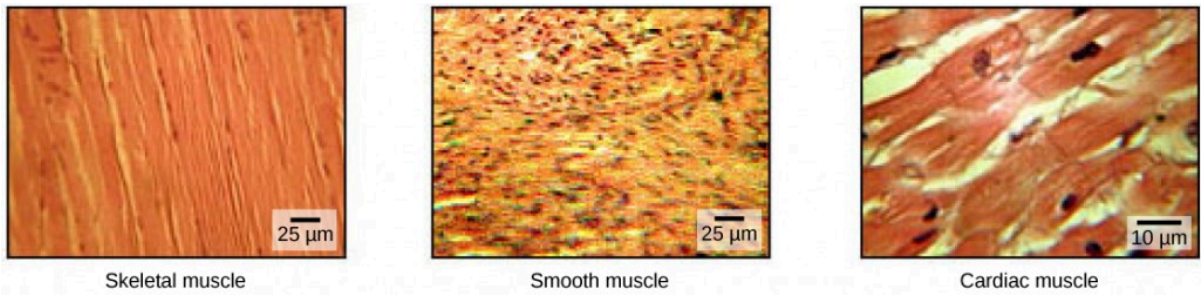


Figure 19.33. The body contains three types of muscle tissue: skeletal muscle, smooth muscle, and cardiac muscle, visualized here using light microscopy. Smooth muscle cells are short, tapered at each end, and have only one plump nucleus in each. Cardiac muscle cells are branched and striated, but short. The cytoplasm may branch, and they have one nucleus in the center of the cell. (credit: modification of work by NCI, NIH; scale-bar data from Matt Russell)

Skeletal muscle tissue forms skeletal muscles, which attach to bones or skin and control locomotion and any movement that can be consciously controlled. Because it can be controlled by thought, skeletal muscle is also called voluntary muscle. Skeletal muscles are long and cylindrical in appearance; when viewed under a microscope, skeletal muscle tissue has a striped or striated appearance. The striations are caused by the regular arrangement of contractile proteins (actin and myosin). **Actin** is a globular contractile protein that interacts with **myosin** for muscle contraction. Skeletal muscle also has multiple nuclei present in a single cell.

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

Cardiac muscle tissue is only found in the heart, and cardiac contractions pump blood throughout the body and maintain blood pressure. Like skeletal muscle, cardiac muscle is striated, but unlike skeletal

muscle, cardiac muscle cannot be consciously controlled and is called involuntary muscle. It has one nucleus per cell, is branched, and is distinguished by the presence of intercalated disks.

Skeletal Muscle Fiber Structure

Each skeletal muscle fiber is a skeletal muscle cell. These cells are incredibly large, with diameters of up to 100 µm and lengths of up to 30 cm. The plasma membrane of a skeletal muscle fiber is called the **sarcolemma**. The sarcolemma is the site of action potential conduction, which triggers muscle contraction. Within each muscle fiber are **myofibrils**—long cylindrical structures that lie parallel to the muscle fiber. Myofibrils run the entire length of the muscle fiber, and because they are only approximately 1.2 µm in diameter, hundreds to thousands can be found inside one muscle fiber. They attach to the sarcolemma at their ends, so that as myofibrils shorten, the entire muscle cell contracts (Figure 19.34).

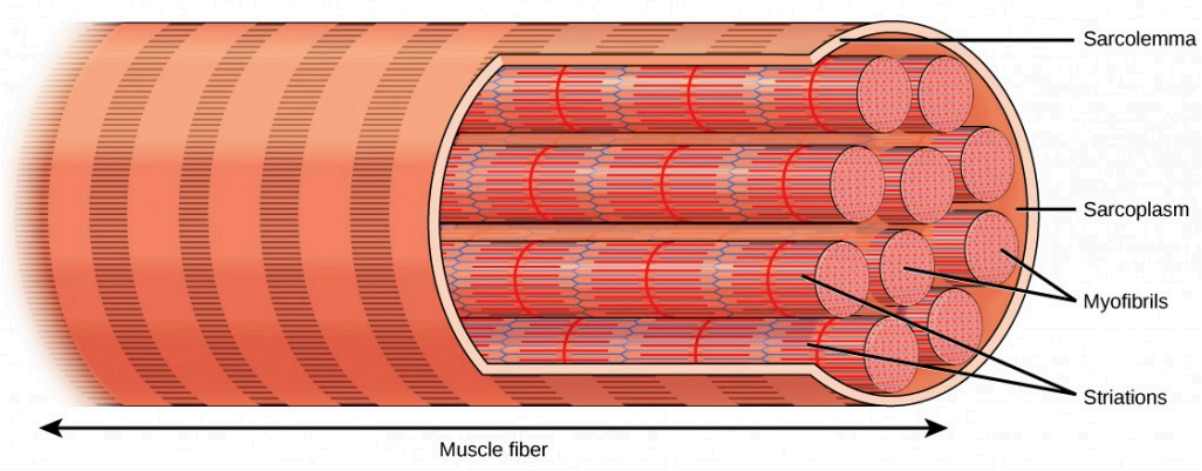


Figure 19.34. A skeletal muscle cell is surrounded by a plasma membrane called the sarcolemma with a cytoplasm called the sarcoplasm. A muscle fiber is composed of many fibrils, packaged into orderly units.

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 19.35).

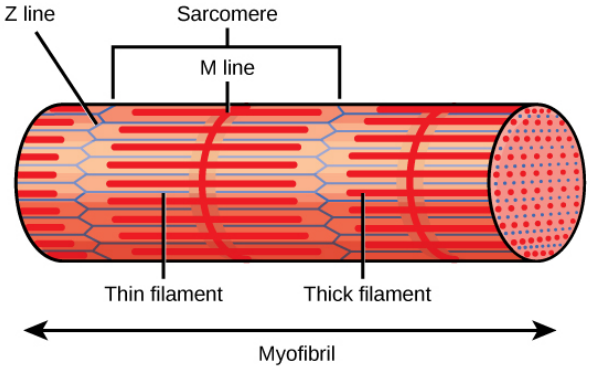


Figure 19.35.
A sarcomere is the region from one Z line to the next Z line. Many sarcomeres are present in a myofibril, resulting in the striation pattern characteristic of skeletal muscle.

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 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.

Concept in Action

View this animation showing the organization of muscle fibers.

Sliding Filament Model of Contraction

For a muscle cell to contract, the sarcomere must shorten. However, thick and thin filaments—the components of sarcomeres—do not shorten. Instead, they slide by one another, causing the sarcomere to shorten while the filaments remain the same length. The sliding filament theory of muscle contraction was developed to fit the differences observed in the named bands on the sarcomere at different degrees of muscle contraction and relaxation. The mechanism of contraction is the binding of myosin to actin, forming cross-bridges that generate filament movement (Figure 19.36).

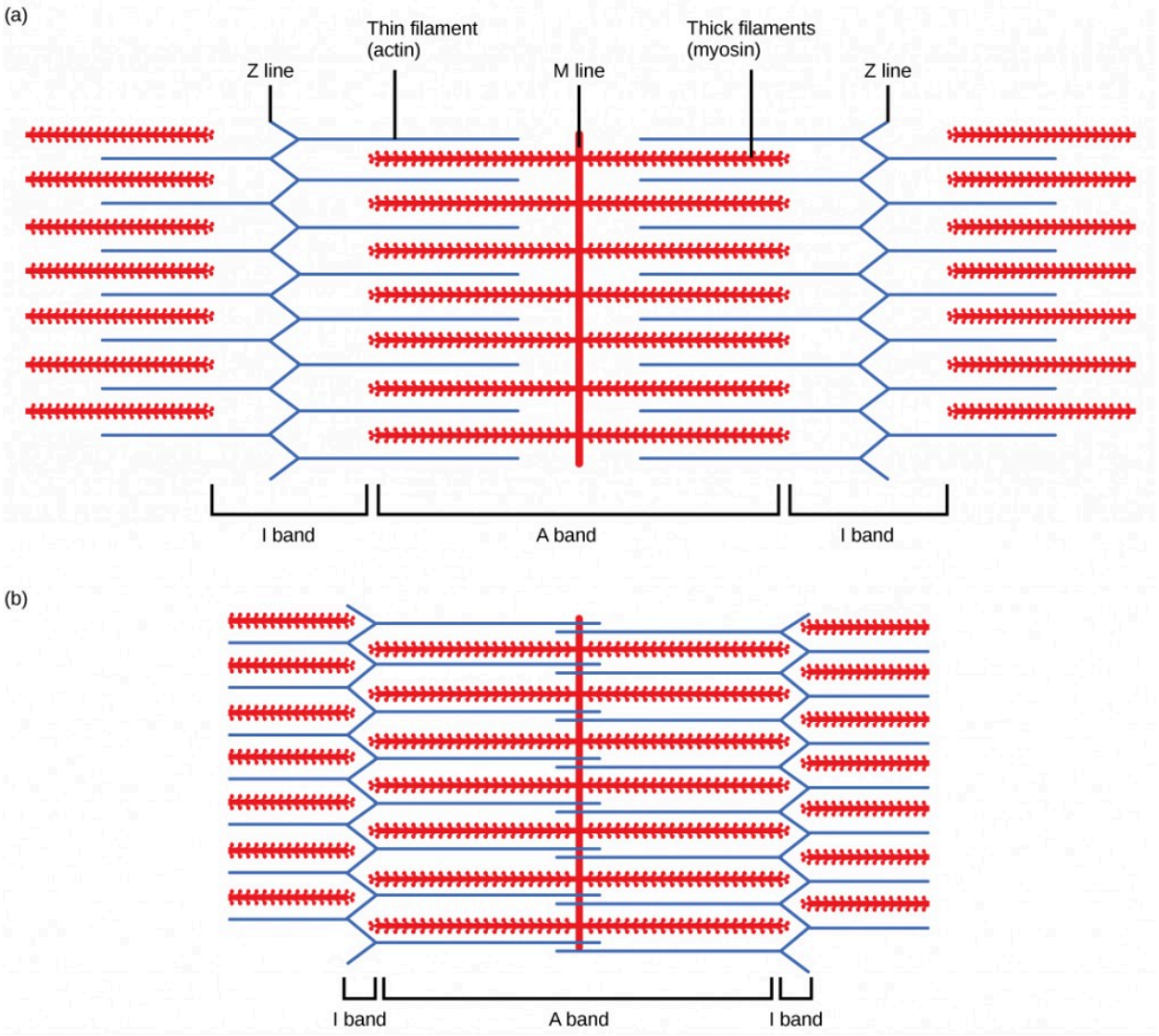


Figure 19.36.
When (a) a sarcomere (b) contracts, the Z lines move closer together and the I band gets smaller. The A band stays the same width and, at full contraction, the thin filaments overlap.

When a sarcomere shortens, some regions shorten whereas others stay the same length. A sarcomere is defined as the distance between two consecutive Z discs or Z lines; when a muscle contracts, the distance between the Z discs is reduced. The H zone—the central region of the A zone—contains only thick filaments and is shortened during contraction. The I band contains only thin filaments and also

shortens. The A band does not shorten—it remains the same length—but A bands of different sarcomeres move closer together during contraction, eventually disappearing. Thin filaments are pulled by the thick filaments toward the center of the sarcomere until the Z discs approach the thick filaments. The zone of overlap, in which thin filaments and thick filaments occupy the same area, increases as the thin filaments move inward.

ATP and Muscle Contraction

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

ATP binding causes myosin to release actin, allowing actin and myosin to detach from each other. After this happens, the newly bound ATP is converted to ADP and inorganic phosphate, 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 19.37).

Concept in Action

Watch this video explaining how a muscle contraction is signaled.

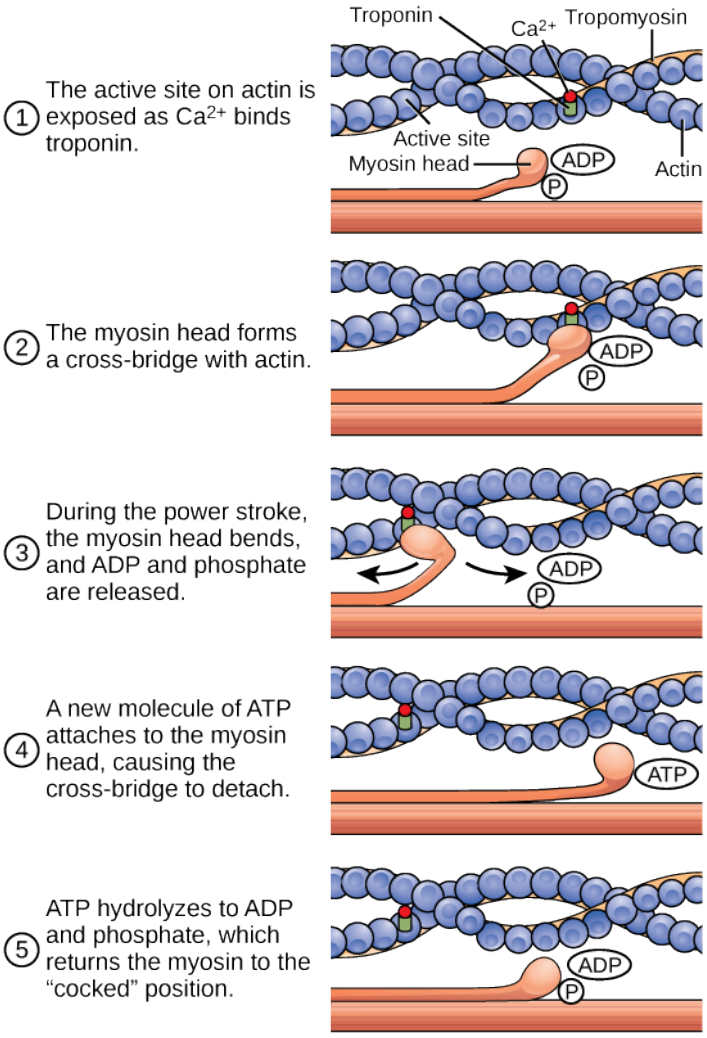


Figure 19.37. The cross-bridge muscle contraction cycle, which is triggered by Ca2+ binding to the actin active site, is shown. With each contraction cycle, actin moves relative to myosin.

Which of the following statements about muscle contraction is true?

1. The power stroke occurs when ATP is hydrolyzed to ADP and phosphate.
2. The power stroke occurs when ADP and phosphate dissociate from the myosin head.
3. The power stroke occurs when ADP and phosphate dissociate from the actin active site.
4. The power stroke occurs when Ca^{2+} binds the calcium head.

Concept in Action

View this animation of the cross-bridge muscle contraction.

Regulatory Proteins

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

To enable a muscle contraction, tropomyosin must change conformation, uncovering the myosin-binding site on an actin molecule and allowing cross-bridge formation. This can only happen in the presence of calcium, which is kept at extremely low concentrations in the sarcoplasm. If present, calcium ions bind to troponin, causing conformational changes in troponin that allow tropomyosin to move away from the myosin binding sites on actin. Once the tropomyosin is removed, a cross-bridge can form between actin and myosin, triggering contraction. Cross-bridge cycling continues until 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.

The ability of cells to communicate electrically requires that the cells expend energy to create an electrical gradient across their cell membranes. This charge gradient is carried by ions, which are differentially distributed across the membrane. Each ion exerts an electrical influence and a concentration influence. Just as milk will eventually mix with coffee without the need to stir, ions also

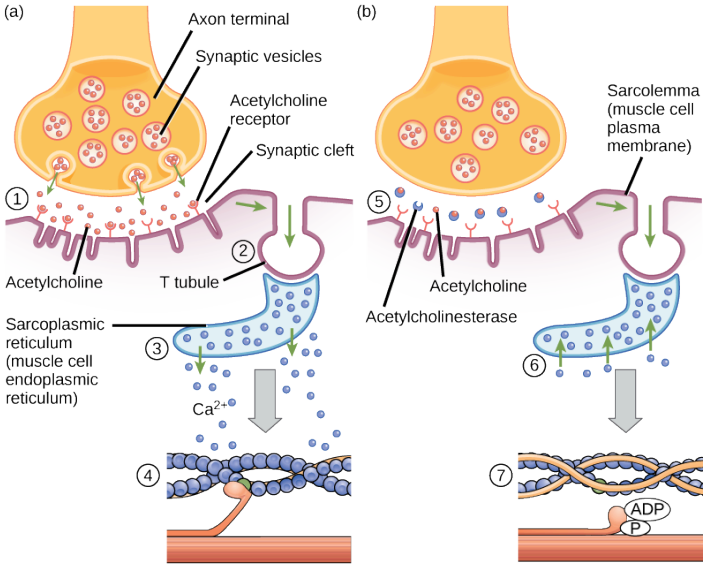
distribute themselves evenly, if they are permitted to do so. In this case, they are not permitted to return to an evenly mixed state.

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

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

Acetylcholine (ACh) is a neurotransmitter released by motor neurons that binds to receptors in the motor end plate. Neurotransmitter release occurs when an action potential travels down the motor neuron’s axon, resulting in altered permeability of the synaptic terminal membrane and an influx of calcium. The Ca^{2+} ions allow synaptic vesicles to move to and bind with the presynaptic membrane (on the neuron), and release neurotransmitter from the vesicles into the synaptic cleft. Once released by the synaptic terminal, ACh diffuses across the synaptic cleft to the motor end plate, where it binds with ACh receptors. As a neurotransmitter binds, these ion channels open, and Na^+ ions cross the membrane into the muscle cell. This reduces the voltage difference between the inside and outside of the cell, which is called depolarization. As ACh binds at the motor end plate, this depolarization is called an end-plate potential. The depolarization then spreads along the sarcolemma, creating an action potential as sodium channels adjacent to the initial depolarization site sense the change in voltage and open. The action potential moves across the entire cell, creating a wave of depolarization.

ACh is broken down by the enzyme **acetylcholinesterase** (AChE) into acetyl and choline. AChE resides in the synaptic cleft, breaking down ACh so that it does not remain bound to ACh receptors, which would cause unwanted extended muscle contraction (Figure 19.38).



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 19.38. This diagram shows excitation-contraction coupling in a skeletal muscle contraction. The sarcoplasmic reticulum is a specialized endoplasmic reticulum found in muscle cells.

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

After depolarization, the membrane returns to its resting state. This is called repolarization, during which voltage-gated sodium channels close. Potassium channels continue at 90% conductance. Because the plasma membrane sodium–potassium ATPase always transports ions, the resting state (negatively charged inside relative to the outside) is restored. The period immediately following the transmission of an impulse in a nerve or muscle, in which a neuron or muscle cell regains its ability to transmit another impulse, is called the refractory period. During the refractory period, the membrane cannot generate another action potential. . The refractory period allows the voltage-sensitive ion channels to return to their resting configurations. The sodium potassium ATPase continually moves Na^+ back out of the cell and K^+ back into the cell, and the K^+ leaks out leaving negative charge behind. Very quickly, the membrane repolarizes, so that it can again be depolarized.

Control of Muscle Tension

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

produced depends on the cross-sectional area of the muscle fiber and the frequency of neural stimulation.

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

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

The primary variable determining force production is the number of myofibers within the muscle that receive an action potential from the neuron that controls that fiber. When using the biceps to pick up a pencil, the motor cortex of the brain only signals a few neurons of the biceps, and only a few myofibers respond. In vertebrates, each myofiber responds fully if stimulated. When picking up a piano, the motor cortex signals all of the neurons in the biceps and every myofiber participates. This is close to the maximum force the muscle can produce. As mentioned above, increasing the frequency of action potentials (the number of signals per second) can increase the force a bit more, because the tropomyosin is flooded with calcium.

Summary

The body contains three types of muscle tissue: skeletal muscle, cardiac muscle, and smooth muscle. Skeleton muscle tissue is composed of sarcomeres, the functional units of muscle tissue. Muscle contraction occurs when sarcomeres shorten, as thick and thin filaments slide past each other, which is called the sliding filament model of muscle contraction. ATP provides the energy for cross-bridge formation and filament sliding. Regulatory proteins, such as troponin and tropomyosin, control cross-bridge formation. Excitation–contraction coupling transduces the electrical signal of the neuron, via acetylcholine, to an electrical signal on the muscle membrane, which initiates force production. The number of muscle fibers contracting determines how much force the whole muscle produces.

Exercises

1. Which of the following statements about muscle contraction is true?

1. The power stroke occurs when ATP is hydrolyzed to ADP and phosphate.
2. The power stroke occurs when ADP and phosphate dissociate from the myosin head.
3. The power stroke occurs when ADP and phosphate dissociate from the actin active site.
4. The power stroke occurs when Ca^{2+} binds the calcium head.

2. The deadly nerve gas Sarin irreversibly inhibits acetycholinesterase. What effect would Sarin have on muscle contraction?
3. In relaxed muscle, the myosin-binding site on actin is blocked by _____.
    1. titin
    2. troponin
    3. myglobin
    4. tropomyosin
4. The cell membrane of a muscle fiber is called a _____.
    1. myofibril
    2. sarcolemma
    3. sarcoplasm
    4. myofilament
5. The muscle relaxes if no new nerve signal arrives. However the neurotransmitter from the previous stimulation is still present in the synapse. The activity of _____ helps to remove this neurotransmitter.
    1. myosin
    2. action potential
    3. tropomyosin
    4. acetylcholinesterase
6. The ability of a muscle to generate tension immediately after stimulation is dependent on:
    1. myosin interaction with the M line
    2. overlap of myosin and actin
    3. actin attachments to the Z line
    4. none of the above
7. How would muscle contractions be affected if ATP was completely depleted in a muscle fiber?
8. What factors contribute to the amount of tension produced in an individual muscle fiber?
9. What effect will low blood calcium have on neurons? What effect will low blood calcium have on skeletal muscles?

Answers

1. B
2. In the presence of Sarin, acetycholine is not removed from the synapse, resulting in continuous stimulation of the muscle plasma membrane. At first, muscle activity is intense and uncontrolled, but the ion gradients dissipate, so electrical signals in the T-tubules are no longer possible. The result is paralysis, leading to death by asphyxiation.
3. D
4. B
5. D
6. D
7. Because ATP is required for myosin to release from actin, muscles would remain rigidly contracted until more ATP was available for the myosin cross-bridge release. This is why dead vertebrates undergo rigor mortis.
8. The cross-sectional area, the length of the muscle fiber at rest, and the frequency of neural stimulation.
9. Neurons will not be able to release neurotransmitter without calcium. Skeletal muscles have calcium stored and don’t need any from the outside.

Glossary

acetylcholinesterase
(AChE) enzyme that breaks down ACh into acetyl and choline

actin
globular contractile protein that interacts with myosin for muscle contraction

motor end plate
sarcolemma of the muscle fiber that interacts with the neuron

myofibril
long cylindrical structures that lie parallel to the muscle fiber

myofilament
small structures that make up myofibrils

myosin
contractile protein that interacts with actin for muscle contraction

osseous tissue
connective tissue that constitutes the endoskeleton

sarcolemma
plasma membrane of a skeletal muscle fiber

sarcomere
functional unit of skeletal muscle

skeletal muscle tissue
forms skeletal muscles, which attach to bones and control locomotion and any movement that can be consciously controlled

spongy bone tissue
forms the inner layer of all bones

thick filament
a group of myosin molecules

thin filament
two polymers of actin wound together along with tropomyosin and troponin

tropomyosin
acts to block myosin binding sites on actin molecules, preventing cross-bridge formation and preventing contraction until a muscle receives a neuron signal

troponin
binds to tropomyosin and helps to position it on the actin molecule, and also binds calcium ions

ulna
bone located on the medial aspect (pinky-finger side) of the forearm

vertebral column
(also, spine) surrounds and protects the spinal cord, supports the head, and acts as an attachment point for ribs and muscles of the back and neck

Chapter 20. The Respiratory System

Figure 20.1.
Lungs, which appear as nearly transparent tissue surrounding the heart in this X-ray of a dog (left), are the central organs of the respiratory system. The left lung is smaller than the right lung to accommodate space for the heart. A dog's nose (right) has a slit on the side of each nostril. When tracking a scent, the slits open, blocking the front of the nostrils. This allows the dog to exhale though the now-open area on the side of the nostrils without losing the scent that is being followed. (credit a: modification of work by Geoff Stearns; credit b: modification of work by Cory Zanker)

Introduction

Breathing is an involuntary event. How often a breath is taken and how much air is inhaled or exhaled are tightly regulated by the respiratory center in the brain. Humans, when they aren't exerting themselves, breathe approximately 15 times per minute on average. Canines, like the dog in Figure 20.1, have a respiratory rate of about 15–30 breaths per minute. With every inhalation, air fills the lungs, and with every exhalation, air rushes back out. That air is doing more than just inflating and deflating the lungs in the chest cavity. The air contains oxygen that crosses the lung tissue, enters the bloodstream, and travels to organs and tissues. Oxygen (O_2) enters the cells where it is used for metabolic reactions that produce ATP, a high-energy compound. At the same time, these reactions release carbon dioxide (CO_2) as a by-product. CO_2 is toxic and must be eliminated. Carbon dioxide exits the cells, enters the bloodstream, travels back to the lungs, and is expired out of the body during exhalation.

20.1 Systems of Gas Exchange

Learning Objectives

By the end of this section, you will be able to:

- Describe the passage of air from the outside environment to the lungs
- Explain how the lungs are protected from particulate matter

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

All aerobic organisms require oxygen to carry out their metabolic functions. Along the evolutionary tree, different organisms have devised different means of obtaining oxygen from the surrounding atmosphere. The environment in which the animal lives greatly determines how an animal respires. The complexity of the respiratory system is correlated with the size of the organism. As animal size increases, diffusion distances increase and the ratio of surface area to volume drops. In unicellular organisms, diffusion across the cell membrane is sufficient for supplying oxygen to the cell (Figure 20.2). Diffusion is a slow, passive transport process. In order for diffusion to be a feasible means of providing oxygen to the cell, the rate of oxygen uptake must match the rate of diffusion across the membrane. In other words, if the cell were very large or thick, diffusion would not be able to provide oxygen quickly enough to the inside of the cell. Therefore, dependence on diffusion as a means of obtaining oxygen and removing carbon dioxide remains feasible only for small organisms or those with highly-flattened bodies, such as many flatworms (Platyhelminthes). Larger organisms had to evolve specialized respiratory tissues, such as gills, lungs, and respiratory passages accompanied by a complex circulatory systems, to transport oxygen throughout their entire body.

Figure 20.2. The cell of the unicellular algae Ventricaria ventricosa is one of the largest known, reaching one to five centimeters in diameter. Like all single-celled organisms, V. ventricosa exchanges gases across the cell membrane.

Direct Diffusion

For small multicellular organisms, diffusion across the outer membrane is sufficient to meet their oxygen needs. Gas exchange by direct diffusion across surface membranes is efficient for organisms less than 1 mm in diameter. In simple organisms, such as cnidarians and flatworms, every cell in the body is close to the external environment. Their cells are kept moist and gases diffuse quickly via direct diffusion. Flatworms are small, literally flat worms, which ‘breathe’ through diffusion across the outer membrane (Figure 20.3). The flat shape of these organisms increases the surface area for diffusion, ensuring that each cell within the body is close to the outer membrane surface and has access to oxygen. If the flatworm had a cylindrical body, then the cells in the center would not be able to get oxygen.

Figure 20.3. 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 cell membranes.

Organisms that live in water need to obtain oxygen from the water. Oxygen dissolves in water but at a lower concentration than in the atmosphere. The atmosphere has roughly 21 percent oxygen. In water, the oxygen concentration is much smaller than that. Fish and many other aquatic organisms have evolved gills to take up the dissolved oxygen from water (Figure 20.4). Gills are thin tissue filaments that are highly branched and folded. When water passes over the gills, the dissolved oxygen in water rapidly diffuses across the gills into the bloodstream. The circulatory system can then carry the oxygenated blood to the other parts of the body. In animals that contain coelomic fluid instead of blood, oxygen diffuses across the gill surfaces into the coelomic fluid. Gills are found in mollusks, annelids, and crustaceans.

Figure 20.4.
This common carp, like many other aquatic organisms, has gills that allow it to obtain oxygen from water. (credit: "Guitardude012"/Wikimedia Commons)

The folded surfaces of the gills provide a large surface area to ensure that the fish gets sufficient oxygen. Diffusion is a process in which material travels from regions of high concentration to low concentration until equilibrium is reached. In this case, blood with a low concentration of oxygen molecules circulates through the gills. The concentration of oxygen molecules in water is higher than the concentration of oxygen molecules in gills. As a result, oxygen molecules diffuse from water (high concentration) to blood (low concentration), as shown in Figure 20.5. Similarly, carbon dioxide molecules in the blood diffuse from the blood (high concentration) to water (low concentration).

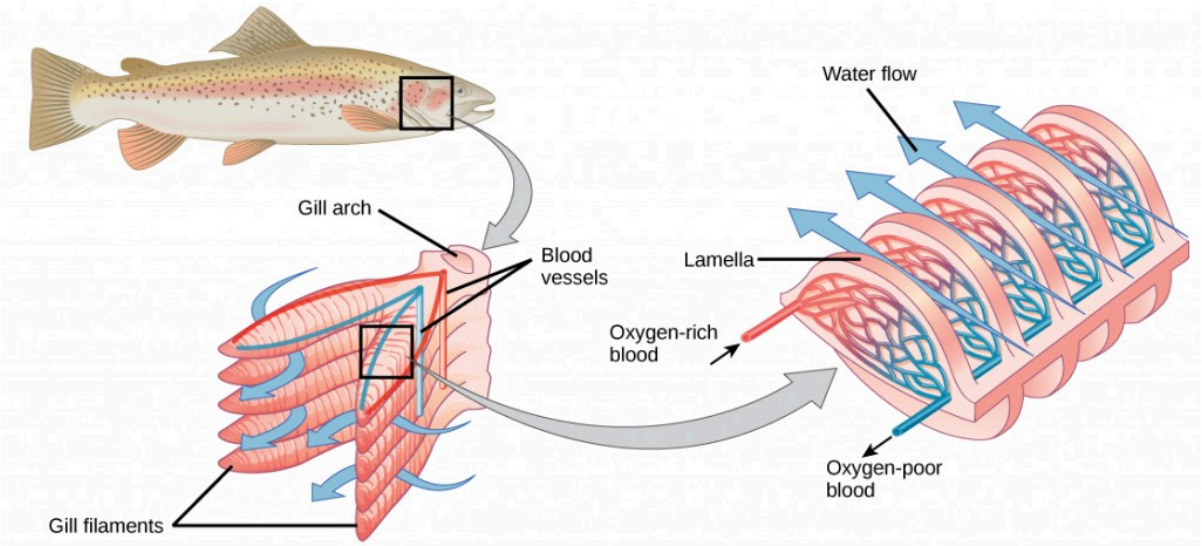


Figure 20.5. As water flows over the gills, oxygen is transferred to blood via the veins. (credit "fish": modification of work by Duane Raver, NOAA)

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. The tubes in the tracheal system are made of a polymeric material called chitin.

Insect bodies have openings, called spiracles, along the thorax and abdomen. These openings connect to the tubular network, allowing oxygen to pass into the body (Figure 20.6) 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.

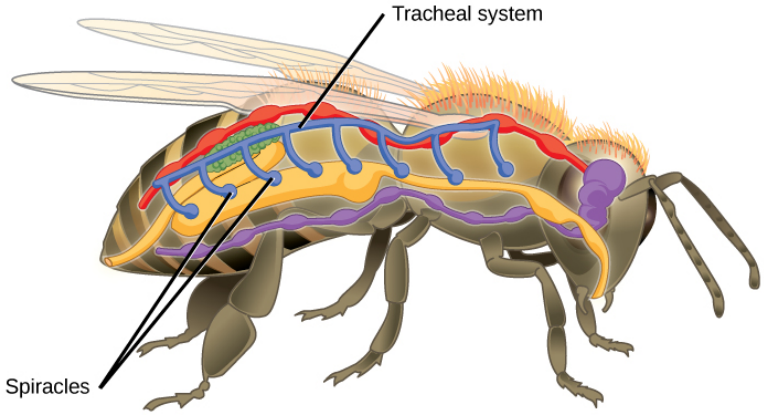


Figure 20.6. Insects perform respiration via a tracheal system.

Mammalian Systems

In mammals, pulmonary ventilation occurs via inhalation (breathing). During inhalation, air enters the

body through the **nasal cavity** located just inside the nose (Figure 20.7). 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.

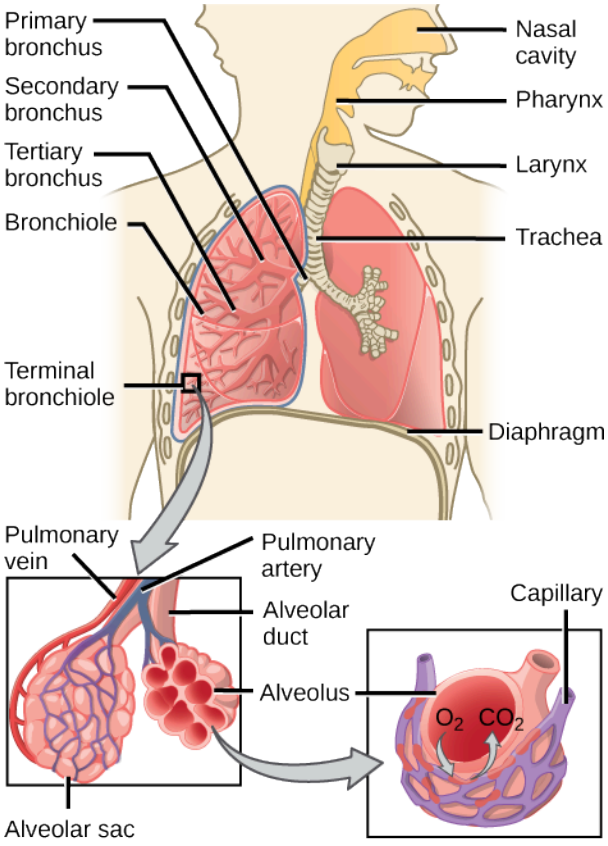


Figure 20.7. Air enters the respiratory system through the nasal cavity and pharynx, and then passes through the trachea and into the bronchi, which bring air into the lungs. (credit: modification of work by NCI)

Which of the following statements about the mammalian respiratory system is false?

1. When we breathe in, air travels from the pharynx to the trachea.
2. The bronchioles branch into bronchi.
3. Alveolar ducts connect to alveolar sacs.
4. Gas exchange between the lung and blood takes place in the alveolus.

From the nasal cavity, air passes through the **pharynx** (throat) and the **larynx** (voice box), as it makes its way to the **trachea** (Figure 20.7). 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. It is made of incomplete rings of hyaline cartilage and smooth muscle (Figure 20.8). 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.

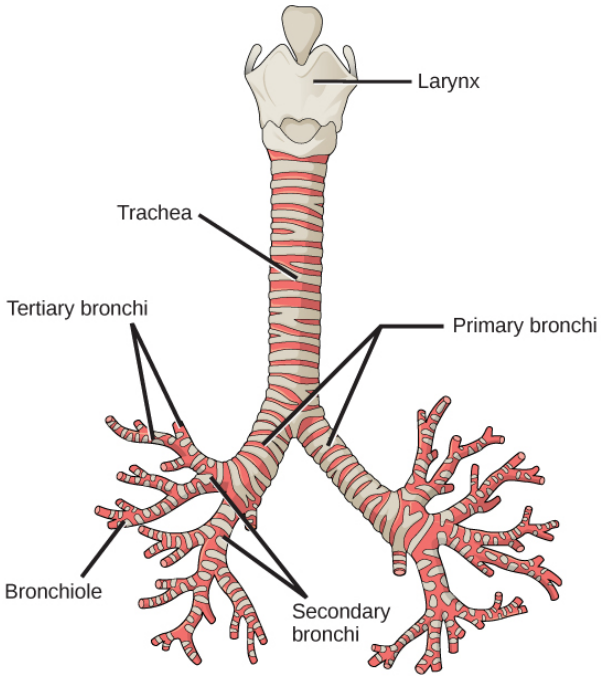


Figure 20.8.
The trachea and bronchi are made of incomplete rings of cartilage. (credit: modification of work by Gray's Anatomy)

Lungs: Bronchi and Alveoli

The end of the trachea bifurcates (divides) to the right and left lungs. The lungs are not identical. The right lung is larger and contains three lobes, whereas the smaller left lung contains two lobes (Figure 20.9). The muscular **diaphragm,** which facilitates breathing, is inferior (below) to the lungs and marks the end of the thoracic cavity.

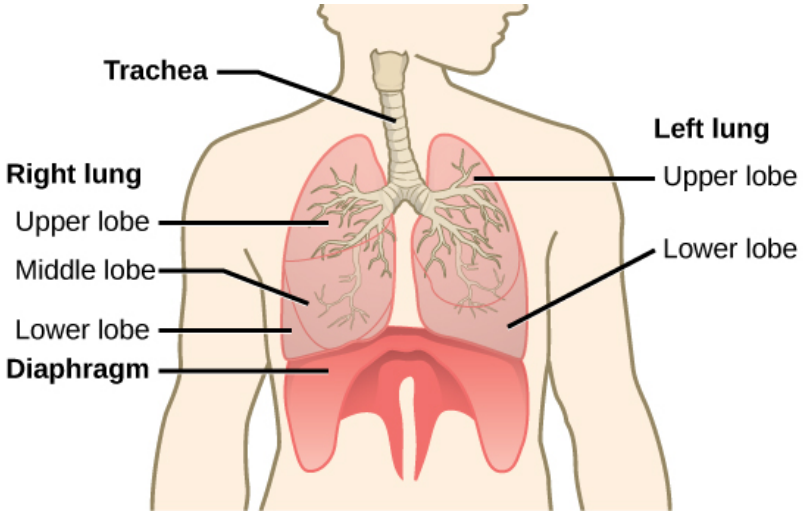


Figure 20.9. The trachea bifurcates into the right and left bronchi in the lungs. The right lung is made of three lobes and is larger. To accommodate the heart, the left lung is smaller and has only two lobes.

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

The **terminal bronchioles** subdivide into microscopic branches called respiratory bronchioles. The respiratory bronchioles subdivide into several alveolar ducts. Numerous alveoli and alveolar sacs surround the alveolar ducts. The alveolar sacs resemble bunches of grapes tethered to the end of the bronchioles (Figure 20.10). In the acinar region, the **alveolar ducts** are attached to the end of each bronchiole. At the end of each duct are approximately 100 **alveolar sacs,** each containing 20 to 30 **alveoli** that are 200 to 300 microns in diameter. Gas exchange occurs only in alveoli. Alveoli are made of thin-walled parenchymal cells, typically one-cell thick, that look like tiny bubbles within the sacs. Alveoli are in direct contact with capillaries (one-cell thick) of the circulatory system. Such intimate contact ensures that oxygen will diffuse from alveoli into the blood and be distributed to the cells of the body. In addition, the carbon dioxide that was produced by cells as a waste product will diffuse from the blood into alveoli to be exhaled. The anatomical arrangement of capillaries and alveoli emphasizes the structural and functional relationship of the respiratory and circulatory systems. Because there are so many alveoli (~300 million per lung) within each alveolar sac and so many sacs at the end of each alveolar duct, the lungs have a sponge-like consistency. This organization produces a very large surface area that is available for gas exchange. The surface area of alveoli in the lungs is approximately 75 m^2. This large surface area, combined with the thin-walled nature of the alveolar parenchymal cells, allows gases to easily diffuse across the cells.

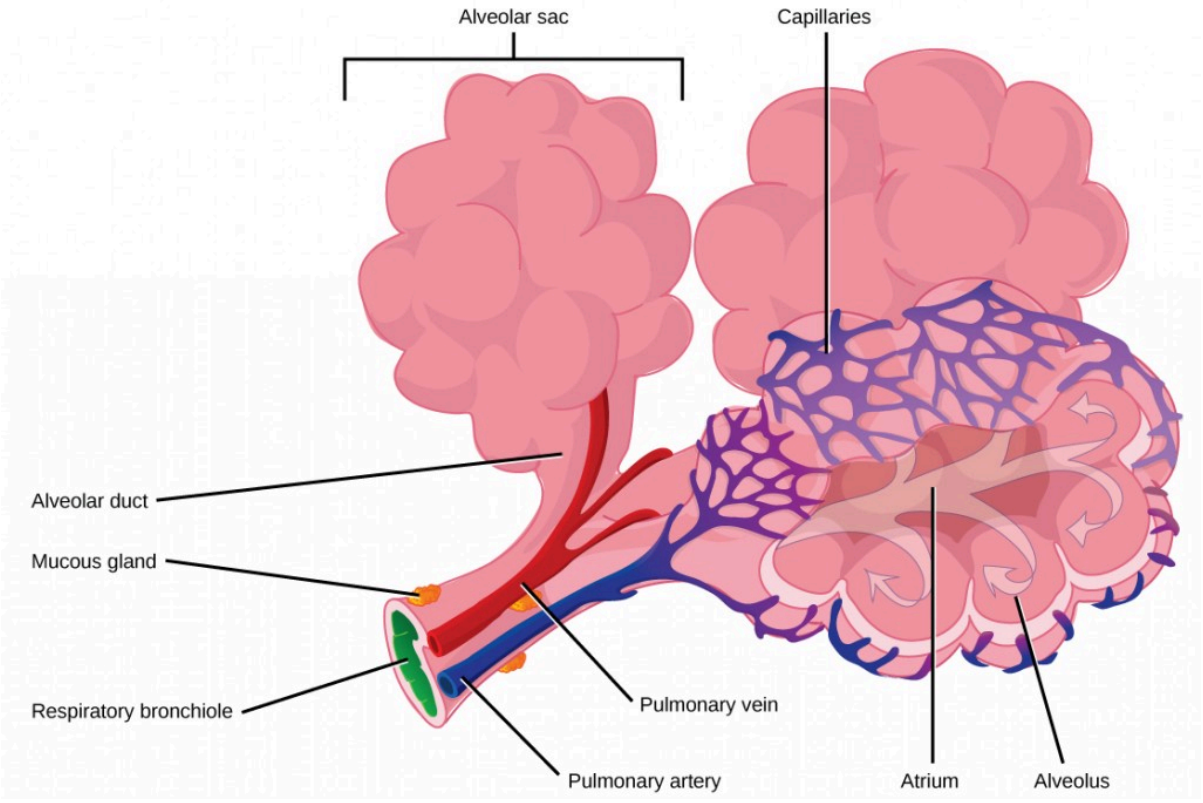


Figure 20.10.
Terminal bronchioles are connected by respiratory bronchioles to alveolar ducts and alveolar sacs. Each alveolar sac contains 20 to 30 spherical alveoli and has the appearance of a bunch of grapes. Air flows into the atrium of the alveolar sac, then circulates into alveoli where gas exchange occurs with the capillaries. Mucous glands secrete mucous into the airways, keeping them moist and flexible. (credit: modification of work by Mariana Ruiz Villareal)

Concept in Action

Watch the following video to review the respiratory system.

Protective Mechanisms

The air that organisms breathe contains **particulate matter** such as dust, dirt, viral particles, and bacteria that can damage the lungs or trigger allergic immune responses. The respiratory system contains several protective mechanisms to avoid problems or tissue damage. In the nasal cavity, hairs and mucus trap small particles, viruses, bacteria, dust, and dirt to prevent their entry.

If particulates do make it beyond the nose, or enter through the mouth, the bronchi and bronchioles of the lungs also contain several protective devices. The lungs produce **mucus**—a sticky substance made of **mucin**, a complex glycoprotein, as well as salts and water—that traps particulates. The bronchi and bronchioles contain cilia, small hair-like projections that line the walls of the bronchi and bronchioles (Figure 20.11). These cilia beat in unison and move mucus and particles out of the bronchi and bronchioles back up to the throat where it is swallowed and eliminated via the esophagus.

In humans, for example, tar and other substances in cigarette smoke destroy or paralyze the cilia, making the removal of particles more difficult. In addition, smoking causes the lungs to produce more mucus, which the damaged cilia are not able to move. This causes a persistent cough, as the lungs try to rid themselves of particulate matter, and makes smokers more susceptible to respiratory ailments.

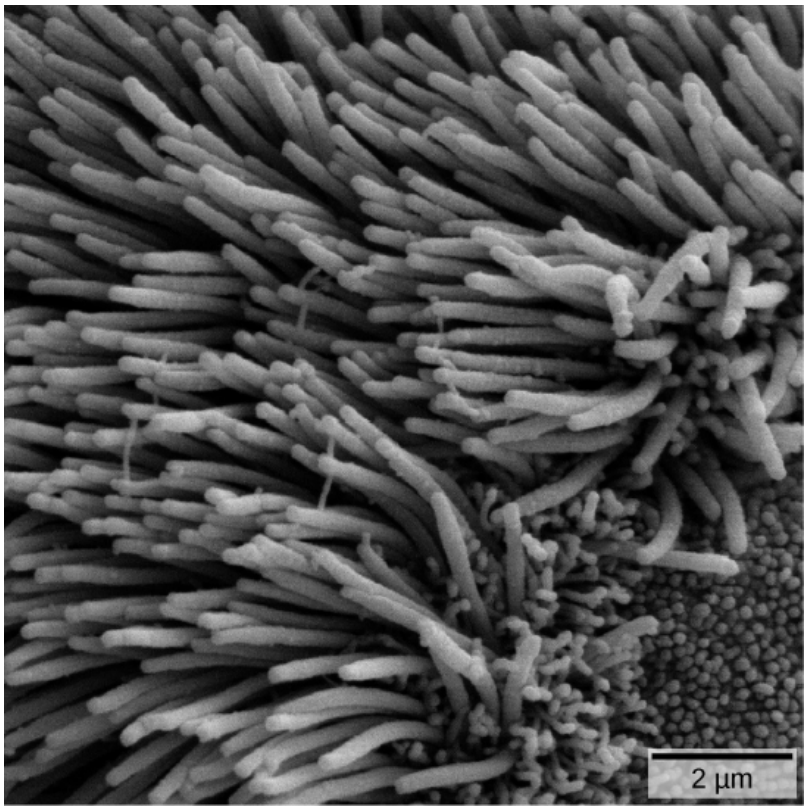


Figure 20.11.
The bronchi and bronchioles contain cilia that help move mucus and other particles out of the lungs. (credit: Louisa Howard, modification of work by Dartmouth Electron Microscope Facility)

Summary

Animal respiratory systems are designed to facilitate gas exchange. In mammals, air is warmed and humidified in the nasal cavity. Air then travels down the pharynx, through the trachea, and into the lungs. In the lungs, air passes through the branching bronchi, reaching the respiratory bronchioles, which house the first site of gas exchange. The respiratory bronchioles open into the alveolar ducts, alveolar sacs, and alveoli. Because there are so many alveoli and alveolar sacs in the lung, the surface area for gas exchange is very large. Several protective mechanisms are in place to prevent damage or infection. These include the hair and mucus in the nasal cavity that trap dust, dirt, and other particulate

matter before they can enter the system. In the lungs, particles are trapped in a mucus layer and transported via cilia up to the esophageal opening at the top of the trachea to be swallowed.

Exercises

1. Which of the following statements about the mammalian respiratory system is false?
    1. When we breathe in, air travels from the pharynx to the trachea.
    2. The bronchioles branch into bronchi.
    3. Alveolar ducts connect to alveolar sacs.
    4. Gas exchange between the lung and blood takes place in the alveolus.
2. The respiratory system _____.
    1. provides body tissues with oxygen
    2. provides body tissues with oxygen and carbon dioxide
    3. establishes how many breaths are taken per minute
    4. provides the body with carbon dioxide
3. Air is warmed and humidified in the nasal passages. This helps to _____.
    1. ward off infection
    2. decrease sensitivity during breathing
    3. prevent damage to the lungs
    4. all of the above
4. Which is the order of airflow during inhalation?
    1. nasal cavity, trachea, larynx, bronchi, bronchioles, alveoli
    2. nasal cavity, larynx, trachea, bronchi, bronchioles, alveoli
    3. nasal cavity, larynx, trachea, bronchioles, bronchi, alveoli
    4. nasal cavity, trachea, larynx, bronchi, bronchioles, alveoli
5. Describe the function of these terms and describe where they are located: main bronchus, trachea, alveoli, and acinus.
6. How does the structure of alveoli maximize gas exchange?

Answers

1. B
2. A
3. C
4. B
5. The main bronchus is the conduit in the lung that funnels air to the airways where gas exchange occurs. The main bronchus attaches the lungs to the very end of the trachea where it bifurcates.

The trachea is the cartilaginous structure that extends from the pharynx to the primary bronchi. It serves to funnel air to the lungs. The alveoli are the sites of gas exchange; they are located at the terminal regions of the lung and are attached to the respiratory bronchioles. The acinus is the structure in the lung where gas exchange occurs.

6. The sac-like structure of the alveoli increases their surface area. In addition, the alveoli are made of thin-walled parenchymal cells. These features allow gases to easily diffuse across the cells.

Glossary

alveolar duct
duct that extends from the terminal bronchiole to the alveolar sac

alveolar sac
structure consisting of two or more alveoli that share a common opening

alveolar ventilation
how much air is in the alveoli

alveolus
(plural: alveoli) (also, air sac) terminal region of the lung where gas exchange occurs

bronchiole
airway that extends from the main tertiary bronchi to the alveolar sac

bronchus
(plural: bronchi) smaller branch of cartilaginous tissue that stems off of the trachea; air is funneled through the bronchi to the region where gas exchange occurs in alveoli

diaphragm
domed-shaped skeletal muscle located under lungs that separates the thoracic cavity from the abdominal cavity

larynx
voice box, a short passageway connecting the pharynx and the trachea

mucin
complex glycoprotein found in mucus

mucus
sticky protein-containing fluid secretion in the lung that traps particulate matter to be expelled from the body

nasal cavity
opening of the respiratory system to the outside environment

particulate matter
small particle such as dust, dirt, viral particles, and bacteria that are in the air

pharynx
throat; a tube that starts in the internal nares and runs partway down the neck, where it opens into the esophagus and the larynx

primary bronchus
(also, main bronchus) region of the airway within the lung that attaches to the trachea and bifurcates to each lung where it branches into secondary bronchi

respiratory bronchiole
terminal portion of the bronchiole tree that is attached to the terminal bronchioles and alveoli ducts, alveolar sacs, and alveoli

respiratory distress syndrome
disease that arises from a deficient amount of surfactant

respiratory quotient (RQ)
ratio of carbon dioxide production to each oxygen molecule consumed

respiratory rate
number of breaths per minute

terminal bronchiole
region of bronchiole that attaches to the respiratory bronchioles

trachea
cartilaginous tube that transports air from the larynx to the primary bronch

20.2 Gas Exchange across Respiratory Surfaces

Learning Objectives

By the end of this section, you will be able to:

- Name and describe lung volumes and capacities
- Understand how gas pressure influences how gases move into and out of the body

The structure of the lung maximizes its surface area to increase gas diffusion. Because of the enormous number of alveoli (approximately 300 million in each human lung), the surface area of the lung is very large (75 m^2). Having such a large surface area increases the amount of gas that can diffuse into and out of the lungs.

Basic Principles of Gas Exchange

Gas exchange during respiration occurs primarily through diffusion. Diffusion is a process in which transport is driven by a concentration gradient. Gas molecules move from a region of high concentration to a region of low concentration. Blood that is low in oxygen concentration and high in carbon dioxide concentration undergoes gas exchange with air in the lungs. The air in the lungs has a higher concentration of oxygen than that of oxygen-depleted blood and a lower concentration of carbon dioxide. This concentration gradient allows for gas exchange during respiration.

Partial pressure is a measure of the concentration of the individual components in a mixture of gases. The total pressure exerted by the mixture is the sum of the partial pressures of the components in the mixture. The rate of diffusion of a gas is proportional to its partial pressure within the total gas mixture. This concept is discussed further in detail below.

Lung Volumes and Capacities

Different animals have different lung capacities based on their activities. Cheetahs have evolved a much higher lung capacity than humans; it helps provide oxygen to all the muscles in the body and allows them to run very fast. Elephants also have a high lung capacity. In this case, it is not because they run fast but because they have a large body and must be able to take up oxygen in accordance with their body size.

Human lung size is determined by genetics, gender, and height. At maximal capacity, an average lung can hold almost six liters of air, but lungs do not usually operate at maximal capacity. Air in the lungs is measured in terms of **lung volumes** and **lung capacities** (Figure 20.12 and Table 20.1). Volume

measures the amount of air for one function (such as inhalation or exhalation). Capacity is any two or more volumes (for example, how much can be inhaled from the end of a maximal exhalation).

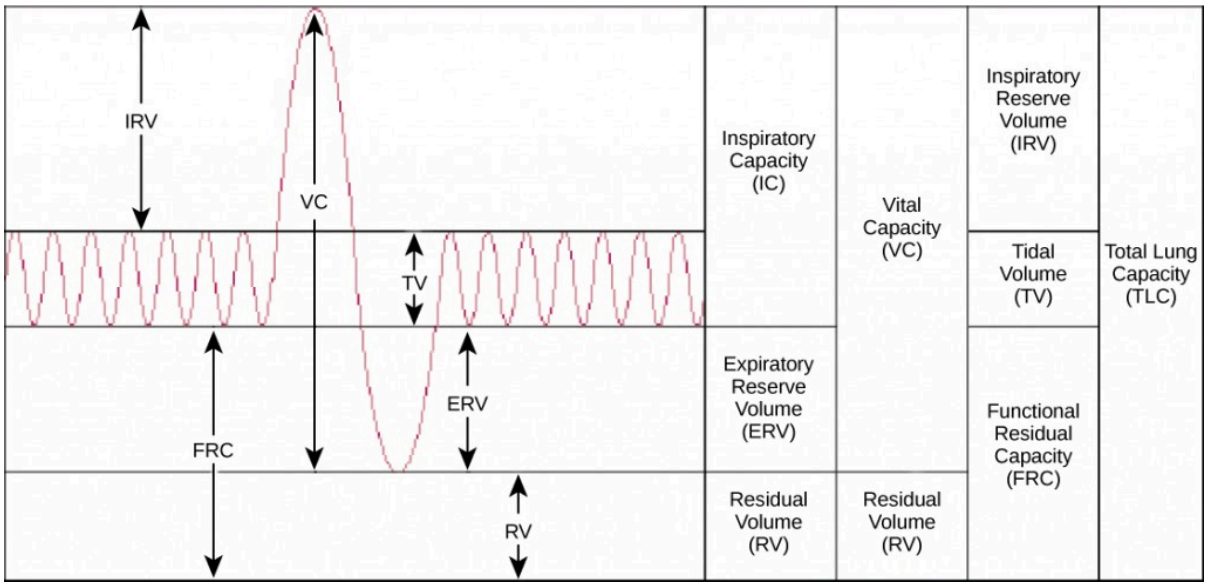


Figure 20.12.
Human lung volumes and capacities are shown. The total lung capacity of the adult male is six liters. Tidal volume is the volume of air inhaled in a single, normal breath. Inspiratory capacity is the amount of air taken in during a deep breath, and residual volume is the amount of air left in the lungs after forceful respiration.

Table 20.1. Lung Volumes and Capacities (Avg Adult Male)

Volume/Capacity	Definition	Volume (liters)	Equations
Tidal volume (TV)	Amount of air inhaled during a normal breath	0.5	–
Expiratory reserve volume (ERV)	Amount of air that can be exhaled after a normal exhalation	1.2	–
Inspiratory reserve volume (IRV)	Amount of air that can be further inhaled after a normal inhalation	3.1	–
Residual volume (RV)	Air left in the lungs after a forced exhalation	1.2	–
Vital capacity (VC)	Maximum amount of air that can be moved in or out of the lungs in a single respiratory cycle	4.8	ERV+TV+IRV
Inspiratory capacity (IC)	Volume of air that can be inhaled in addition to a normal exhalation	3.6	TV+IRV
Functional residual capacity (FRC)	Volume of air remaining after a normal exhalation	2.4	ERV+RV
Total lung capacity (TLC)	Total volume of air in the lungs after a maximal inspiration	6.0	RV+ERV+TV+IRV
Forced expiratory volume (FEV1)	How much air can be forced out of the lungs over a specific time period, usually one second	~4.1 to 5.5	–

The volume in the lung can be divided into four units: tidal volume, expiratory reserve volume, inspiratory reserve volume, and residual volume. **Tidal volume (TV)** measures the amount of air that is inspired and expired during a normal breath. On average, this volume is around one-half liter, which is a little less than the capacity of a 20-ounce drink bottle. The **expiratory reserve volume (ERV)** is the additional amount of air that can be exhaled after a normal exhalation. It is the reserve amount that can be exhaled beyond what is normal. Conversely, the **inspiratory reserve volume (IRV)** is the additional amount of air that can be inhaled after a normal inhalation. The **residual volume (RV)** is the amount of air that is left after expiratory reserve volume is exhaled. The lungs are never completely empty: There is always some air left in the lungs after a maximal exhalation. If this residual volume did not exist and the lungs emptied completely, the lung tissues would stick together and the energy necessary to re-inflate the lung could be too great to overcome. Therefore, there is always some air remaining in the lungs. Residual volume is also important for preventing large fluctuations in respiratory gases (O_2 and CO_2). The residual volume is the only lung volume that cannot be measured directly because it is impossible to completely empty the lung of air. This volume can only be calculated rather than measured.

Capacities are measurements of two or more volumes. The **vital capacity (VC)** measures the maximum amount of air that can be inhaled or exhaled during a respiratory cycle. It is the sum of the expiratory reserve volume, tidal volume, and inspiratory reserve volume. The **inspiratory capacity (IC)** is the amount of air that can be inhaled after the end of a normal expiration. It is, therefore, the sum of the tidal volume and inspiratory reserve volume. The **functional residual capacity (FRC)**

includes the expiratory reserve volume and the residual volume. The FRC measures the amount of additional air that can be exhaled after a normal exhalation. Lastly, the **total lung capacity (TLC)** is a measurement of the total amount of air that the lung can hold. It is the sum of the residual volume, expiratory reserve volume, tidal volume, and inspiratory reserve volume.

Lung volumes are measured by a technique called **spirometry**. An important measurement taken during spirometry is the **forced expiratory volume (FEV)**, which measures how much air can be forced out of the lung over a specific period, usually one second (FEV1). In addition, the forced vital capacity (FVC), which is the total amount of air that can be forcibly exhaled, is measured. The ratio of these values (**FEV1/FVC ratio**) is used to diagnose lung diseases including asthma, emphysema, and fibrosis. If the FEV1/FVC ratio is high, the lungs are not compliant (meaning they are stiff and unable to bend properly), and the patient most likely has lung fibrosis. Patients exhale most of the lung volume very quickly. Conversely, when the FEV1/FVC ratio is low, there is resistance in the lung that is characteristic of asthma. In this instance, it is hard for the patient to get the air out of his or her lungs, and it takes a long time to reach the maximal exhalation volume. In either case, breathing is difficult and complications arise.

Respiratory Therapist

Respiratory therapists or respiratory practitioners evaluate and treat patients with lung and cardiovascular diseases. They work as part of a medical team to develop treatment plans for patients. Respiratory therapists may treat premature babies with underdeveloped lungs, patients with chronic conditions such as asthma, or older patients suffering from lung disease such as emphysema and chronic obstructive pulmonary disease (COPD). They may operate advanced equipment such as compressed gas delivery systems, ventilators, blood gas analyzers, and resuscitators. Specialized programs to become a respiratory therapist generally lead to a bachelor’s degree with a respiratory therapist specialty. Because of a growing aging population, career opportunities as a respiratory therapist are expected to remain strong.

Gas Pressure and Respiration

The respiratory process can be better understood by examining the properties of gases. Gases move freely, but gas particles are constantly hitting the walls of their vessel, thereby producing gas pressure.

Air is a mixture of gases, primarily nitrogen (N_2; 78.6 percent), oxygen (O_2; 20.9 percent), water vapor (H_2O; 0.5 percent), and carbon dioxide (CO_2; 0.04 percent). Each gas component of that mixture exerts a pressure. The pressure for an individual gas in the mixture is the partial pressure of that gas. Approximately 21 percent of atmospheric gas is oxygen. Carbon dioxide, however, is found in relatively small amounts, 0.04 percent. The partial pressure for oxygen is much greater than that of carbon dioxide. The partial pressure of any gas can be calculated by:
(39.1)
$P = (P_{atm}) \times$ (percent content in mixture).

P_{atm}, the atmospheric pressure, is the sum of all of the partial pressures of the atmospheric gases added together,
(39.2)

$P_{atm} = P_{N_2} + P_{O2} + P_{H2O} + P_{CO2} = 760$ mm Hg

× (percent content in mixture).

The pressure of the atmosphere at sea level is 760 mm Hg. Therefore, the partial pressure of oxygen is:
(39.3)
$P_{O_2} = (760 \text{ mm Hg})\ (0.21) = 160 \text{ mm Hg}$

and for carbon dioxide:
(39.4)
$PCO_2 = (760 \text{ mm Hg})\ (0.0004) = 0.3 \text{ mm Hg}.$

At high altitudes, P_{atm} decreases but concentration does not change; the partial pressure decrease is due to the reduction in P_{atm}.

When the air mixture reaches the lung, it has been humidified. The pressure of the water vapor in the lung does not change the pressure of the air, but it must be included in the partial pressure equation. For this calculation, the water pressure (47 mm Hg) is subtracted from the atmospheric pressure:
(39.5)
760 mm Hg − 47 mm Hg = 713 mm Hg

and the partial pressure of oxygen is:
(39.6)
(760 mm Hg − 47 mm Hg) × 0.21 = 150 mm Hg.

These pressures determine the gas exchange, or the flow of gas, in the system. Oxygen and carbon dioxide will flow according to their pressure gradient from high to low. Therefore, understanding the partial pressure of each gas will aid in understanding how gases move in the respiratory system.

Gas Exchange across the Alveoli

In the body, oxygen is used by cells of the body's tissues and carbon dioxide is produced as a waste product. The ratio of carbon dioxide production to oxygen consumption is the **respiratory quotient (RQ)**. RQ varies between 0.7 and 1.0. If just glucose were used to fuel the body, the RQ would equal one. One mole of carbon dioxide would be produced for every mole of oxygen consumed. Glucose, however, is not the only fuel for the body. Protein and fat are also used as fuels for the body. Because of this, less carbon dioxide is produced than oxygen is consumed and the RQ is, on average, about 0.7 for fat and about 0.8 for protein.

The RQ is used to calculate the partial pressure of oxygen in the alveolar spaces within the lung, the **alveolar** P_{O_2} Above, the partial pressure of oxygen in the lungs was calculated to be 150 mm Hg. However, lungs never fully deflate with an exhalation; therefore, the inspired air mixes with this residual air and lowers the partial pressure of oxygen within the alveoli. This means that there is a lower concentration of oxygen in the lungs than is found in the air outside the body. Knowing the RQ, the partial pressure of oxygen in the alveoli can be calculated:

(39.7)

$$\text{alveolar } P_{O_2} = \text{inspired } P_{O_2} - \left(\frac{\text{alveolar } P_{O_2}}{RQ}\right)$$

With an RQ of 0.8 and a PCO_2 in the alveoli of 40 mm Hg, the alveolar PO_2

is equal to:

(39.8)

$$\text{alveolar } P_{O_2} = 150 \text{ mm Hg} - \left(\frac{40 \text{ mm Hg}}{0.8}\right) = \text{mm Hg}.$$

Notice that this pressure is less than the external air. Therefore, the oxygen will flow from the inspired air in the lung (P_{O_2} = 150 mm Hg) into the bloodstream (P_{O_2} = 100 mm Hg)

(Figure 20.13).

In the lungs, oxygen diffuses out of the alveoli and into the capillaries surrounding the alveoli. Oxygen (about 98 percent) binds reversibly to the respiratory pigment hemoglobin found in red blood cells (RBCs). RBCs carry oxygen to the tissues where oxygen dissociates from the hemoglobin and diffuses into the cells of the tissues. More specifically, alveolar P_{O_2} is higher in the alveoli (P_{ALVO2} = 100 mm Hg) than blood P_{O_2} (40 mm Hg) in the capillaries. Because this pressure gradient exists, oxygen diffuses down its pressure gradient, moving out of the alveoli and entering the blood of the capillaries where O_2 binds to hemoglobin. At the same time, alveolar P_{CO2} is lower P_{ALVO2} = 40 mm Hg than blood P_{CO2} = (45 mm Hg). CO_2 diffuses down its pressure gradient, moving out of the capillaries and entering the alveoli.

Oxygen and carbon dioxide move independently of each other; they diffuse down their own pressure gradients. As blood leaves the lungs through the pulmonary veins, the **venous P_{O_2}**= 100 mm Hg, whereas the *venous* P_{CO2} = 40 mm Hg. As blood enters the systemic capillaries, the blood will lose oxygen and gain carbon dioxide because of the pressure difference of the tissues and blood. In systemic capillaries, P_{O_2}= 100 mm Hg, but in the tissue cells, P_{O_2}= 40 mm Hg. This pressure gradient drives the diffusion of oxygen out of the capillaries and into the tissue cells. At the same time, blood P_{CO2}= 40 mm Hg and systemic tissue P_{CO2}= 45 mm Hg. The pressure gradient drives CO_2 out of tissue cells and into the capillaries. The blood returning to the lungs through the pulmonary arteries has a venous P_{O_2}= 40 mm Hg and a P_{CO2}= 45 mm Hg. The blood enters the lung capillaries where the process of exchanging gases between the capillaries and alveoli begins again (Figure 20.13).

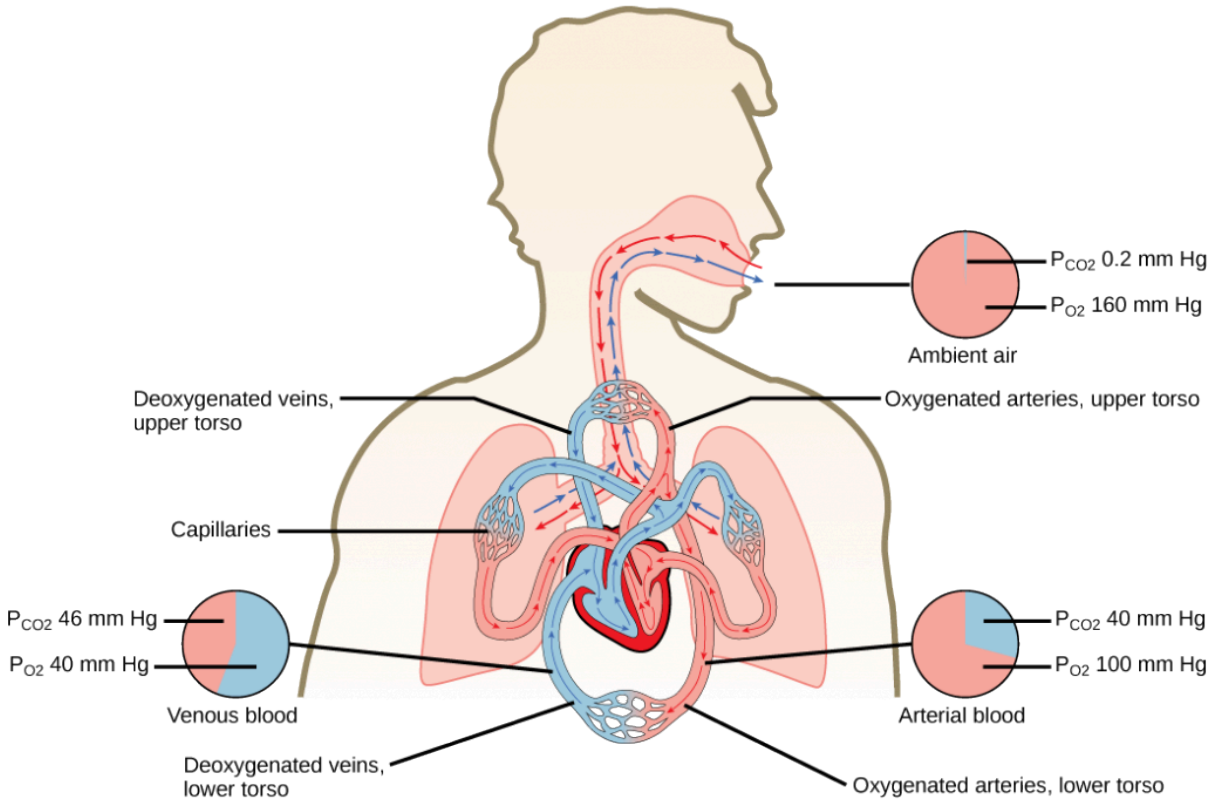


Figure 20.13. The partial pressures of oxygen and carbon dioxide change as blood moves through the body.

Which of the following statements is false?

1. In the tissues, P_{O_2} drops as blood passes from the arteries to the veins, while P_{CO2} increases.
2. Blood travels from the lungs to the heart to body tissues, then back to the heart, then the lungs.
3. Blood travels from the lungs to the heart to body tissues, then back to the lungs, then the heart.
4. P_{O_2} is higher in air than in the lungs.

In short, the change in partial pressure from the alveoli to the capillaries drives the oxygen into the tissues and the carbon dioxide into the blood from the tissues. The blood is then transported to the lungs where differences in pressure in the alveoli result in the movement of carbon dioxide out of the blood into the lungs, and oxygen into the blood.

Concept in Action

Watch this video to learn how to carry out spirometry.

Summary

The lungs can hold a large volume of air, but they are not usually filled to maximal capacity. Lung volume measurements include tidal volume, expiratory reserve volume, inspiratory reserve volume, and residual volume. The sum of these equals the total lung capacity. Gas movement into or out of the lungs is dependent on the pressure of the gas. Air is a mixture of gases; therefore, the partial pressure of each gas can be calculated to determine how the gas will flow in the lung. The difference between the partial pressure of the gas in the air drives oxygen into the tissues and carbon dioxide out of the body.

Exercises

1. Which of the following statements is false?
    1. In the tissues, P_{O2} drops as blood passes from the arteries to the veins, while PC_{O2} increases.
    2. Blood travels from the lungs to the heart to body tissues, then back to the heart, then the lungs.
    3. Blood travels from the lungs to the heart to body tissues, then back to the lungs, then the heart.
    4. P_{O2} is higher in air than in the lungs.
2. The inspiratory reserve volume measures the _____.
    1. amount of air remaining in the lung after a maximal exhalation
    2. amount of air that the lung holds
    3. amount of air the can be further exhaled after a normal breath
    4. amount of air that can be further inhaled after a normal breath
3. Of the following, which does not explain why the partial pressure of oxygen is lower in the lung than in the external air?
    1. Air in the lung is humidified; therefore, water vapor pressure alters the pressure.
    2. Carbon dioxide mixes with oxygen.

3. Oxygen is moved into the blood and is headed to the tissues.
4. Lungs exert a pressure on the air to reduce the oxygen pressure.

4. The total lung capacity is calculated using which of the following formulas?

    1. residual volume + tidal volume + inspiratory reserve volume
    2. residual volume + expiratory reserve volume + inspiratory reserve volume
    3. expiratory reserve volume + tidal volume + inspiratory reserve volume
    4. residual volume + expiratory reserve volume + tidal volume + inspiratory reserve volume

5. What does FEV1/FVC measure? What factors may affect FEV1/FVC?
6. What is the reason for having residual volume in the lung?
7. How can a decrease in the percent of oxygen in the air affect the movement of oxygen in the body?
8. If a patient has increased resistance in his or her lungs, how can this detected by a doctor? What does this mean?

Answers

1. C
2. D
3. D
4. D
5. FEV1/FVC measures the forced expiratory volume in one second in relation to the total forced vital capacity (the total amount of air that is exhaled from the lung from a maximal inhalation). This ratio changes with alterations in lung function that arise from diseases such as fibrosis, asthma, and COPD.
6. If all the air in the lung were exhaled, then opening the alveoli for the next inspiration would be very difficult. This is because the tissues would stick together.
7. Oxygen moves from the lung to the bloodstream to the tissues according to the pressure gradient. This is measured as the partial pressure of oxygen. If the amount of oxygen drops in the inspired air, there would be reduced partial pressure. This would decrease the driving force that moves the oxygen into the blood and into the tissues. P_{O2} is also reduced at high elevations:P_{O2} at high elevations is lower than at sea level because the total atmospheric pressure is less than atmospheric pressure at sea level.
8. A doctor can detect a restrictive disease using spirometry. By detecting the rate at which air can be expelled from the lung, a diagnosis of fibrosis or another restrictive disease can be made.

Glossary

alveolarPO_2
partial pressure of oxygen in the alveoli (usually around 100 mmHg)

expiratory reserve volume (ERV)
amount of additional air that can be exhaled after a normal exhalation

FEV1/FVC ratio
ratio of how much air can be forced out of the lung in one second to the total amount that is forced out of the lung; a measurement of lung function that can be used to detect disease states

forced expiratory volume (FEV)
(also, forced vital capacity) measure of how much air can be forced out of the lung from maximal inspiration over a specific amount of time

functional residual capacity (FRC)
expiratory reserve volume plus residual volume

functional vital capacity (FVC)
amount of air that can be forcibly exhaled after taking the deepest breath possible

inspiratory capacity (IC)
tidal volume plus inspiratory reserve volume

inspiratory reserve volume (IRV)
amount of additional air that can be inspired after a normal inhalation

lung capacity
measurement of two or more lung volumes (how much air can be inhaled from the end of an expiration to maximal capacity)

lung volume
measurement of air for one lung function (normal inhalation or exhalation)

oxygen-carrying capacity
amount of oxygen that can be transported in the blood

partial pressure
amount of pressure exerted by one gas within a mixture of gases

residual volume (RV)
amount of air remaining in the lung after a maximal expiration

respiratory quotient (RQ)
ratio of carbon dioxide production to each oxygen molecule consumed

spirometry
method to measure lung volumes and to diagnose lung diseases

tidal volume (TV)
amount of air that is inspired and expired during normal breathing

venousPCO_2
partial pressure of carbon dioxide in the veins (40 mm Hg in the pulmonary veins)

venousPO_2
partial pressure of oxygen in the veins (100 mm Hg in the pulmonary veins)

20.3 Breathing

Learning Objectives

By the end of this section, you will be able to:

- Describe how the structures of the lungs and thoracic cavity control the mechanics of breathing
- Explain the importance of compliance and resistance in the lungs
- Discuss problems that may arise due to a V/Q mismatch

Mammalian lungs are located in the thoracic cavity where they are surrounded and protected by the rib cage, intercostal muscles, and bound by the chest wall. The bottom of the lungs is contained by the diaphragm, a skeletal muscle that facilitates breathing. Breathing requires the coordination of the lungs, the chest wall, and most importantly, the diaphragm.

Types of Breathing

Amphibians have evolved multiple ways of breathing. Young amphibians, like tadpoles, use gills to breathe, and they don't leave the water. Some amphibians retain gills for life. As the tadpole grows, the gills disappear and lungs grow. These lungs are primitive and not as evolved as mammalian lungs. Adult amphibians are lacking or have a reduced diaphragm, so breathing via lungs is forced. The other means of breathing for amphibians is diffusion across the skin. To aid this diffusion, amphibian skin must remain moist.

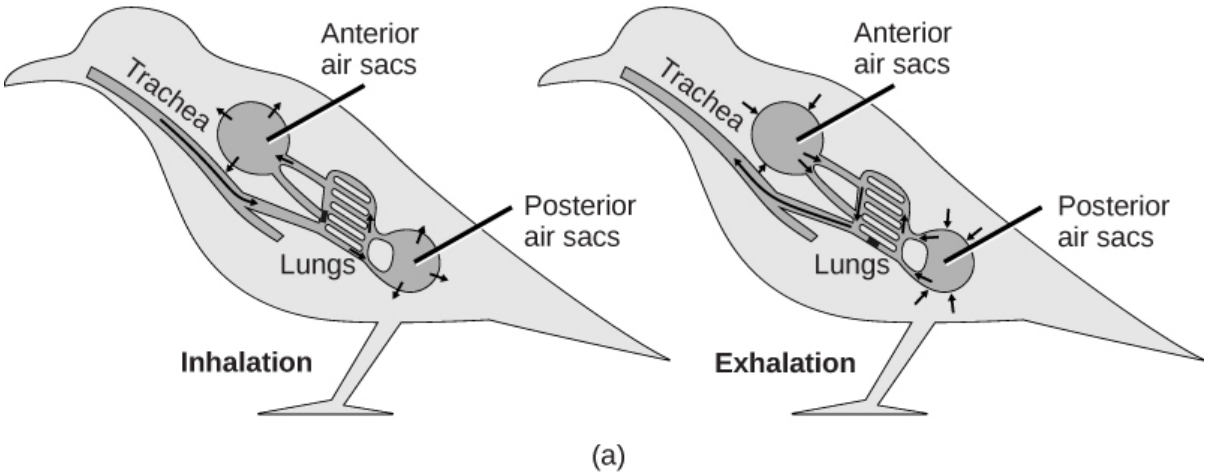
Birds face a unique challenge with respect to breathing: They fly. Flying consumes a great amount of energy; therefore, birds require a lot of oxygen to aid their metabolic processes. Birds have evolved a respiratory system that supplies them with the oxygen needed to enable flying. Similar to mammals, birds have lungs, which are organs specialized for gas exchange. Oxygenated air, taken in during inhalation, diffuses across the surface of the lungs into the bloodstream, and carbon dioxide diffuses from the blood into the lungs and expelled during exhalation. The details of breathing between birds and mammals differ substantially.

In addition to lungs, birds have air sacs inside their body. Air flows in one direction from the posterior air sacs to the lungs and out of the anterior air sacs. The flow of air is in the opposite direction from blood flow, and gas exchange takes place much more efficiently. This type of breathing enables birds to obtain the requisite oxygen, even at higher altitudes where the oxygen concentration is low. This directionality of airflow requires two cycles of air intake and exhalation to completely get the air out of the lungs.

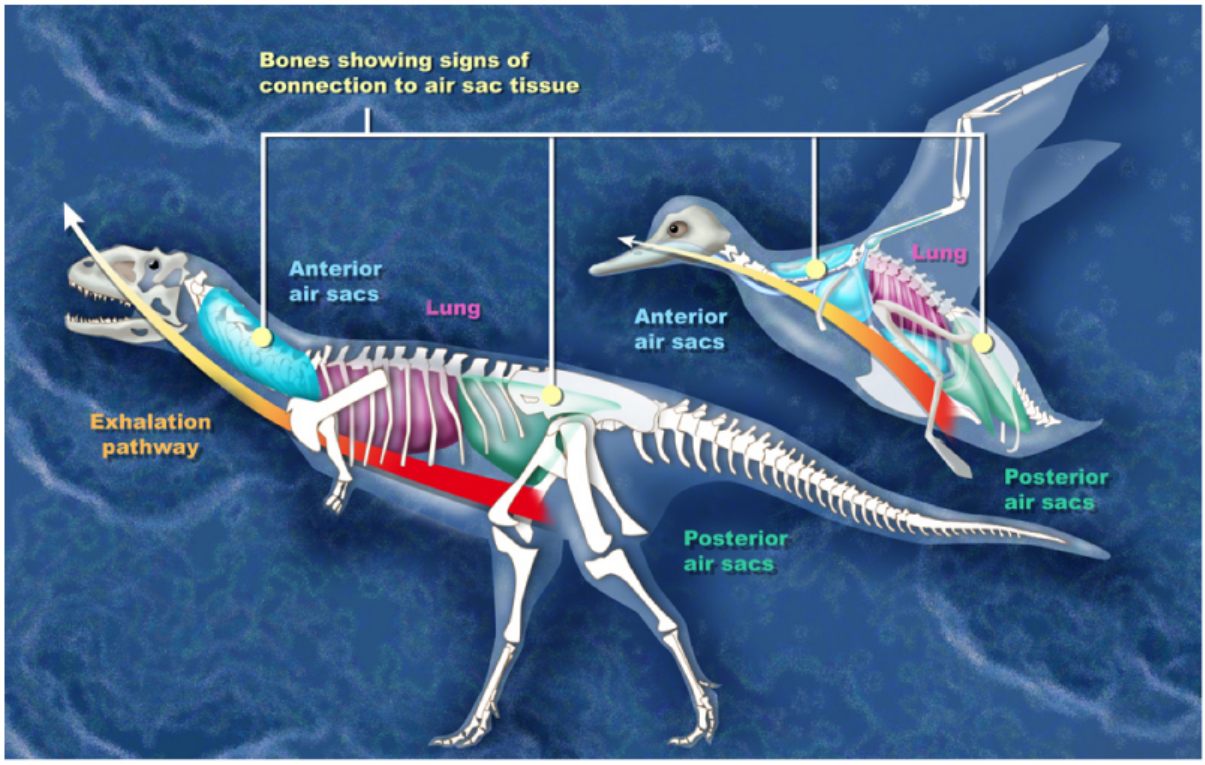
Avian Respiration

Birds have evolved a respiratory system that enables them to fly. Flying is a high-energy process and requires a lot of oxygen. Furthermore, many birds fly in high altitudes where the concentration of oxygen in low. How did birds evolve a respiratory system that is so unique?

Decades of research by paleontologists have shown that birds evolved from therapods, meat-eating dinosaurs (Figure 20.14). In fact, fossil evidence shows that meat-eating dinosaurs that lived more than 100 million years ago had a similar flow-through respiratory system with lungs and air sacs. *Archaeopteryx* and *Xiaotingia*, for example, were flying dinosaurs and are believed to be early precursors of birds.



(a)



(b)

Figure 20.14.
(a) Birds have a flow-through respiratory system in which air flows unidirectionally from the posterior sacs into the lungs, then into the anterior air sacs. The air sacs connect to openings in hollow bones. (b) Dinosaurs, from which birds descended, have similar hollow bones and are believed to have had a similar respiratory system. (credit b: modification of work by Zina Deretsky, National Science Foundation)

Most of us consider that dinosaurs are extinct. However, modern birds are descendants of avian dinosaurs. The respiratory system of modern birds has been evolving for hundreds of millions of years.

All mammals have lungs that are the main organs for breathing. Lung capacity has evolved to support the animal's activities. During inhalation, the lungs expand with air, and oxygen diffuses across the

lung's surface and enters the bloodstream. During exhalation, the lungs expel air and lung volume decreases. In the next few sections, the process of human breathing will be explained.

The Mechanics of Human Breathing

Boyle's Law is the gas law that states that in a closed space, pressure and volume are inversely related. As volume decreases, pressure increases and vice versa (Figure 20.15). The relationship between gas pressure and volume helps to explain the mechanics of breathing.

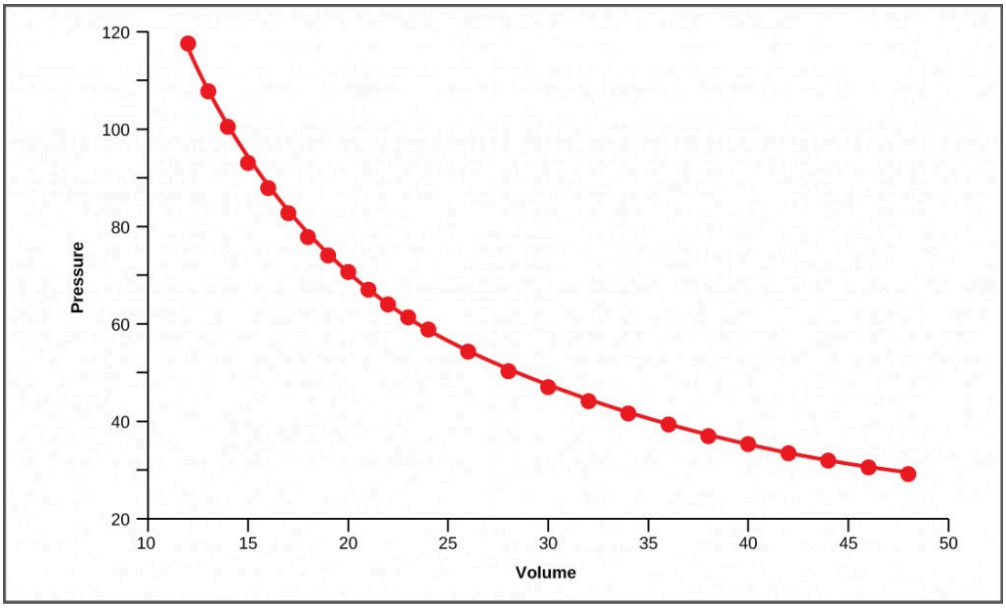


Figure 20.15.
This graph shows data from Boyle's original 1662 experiment, which shows that pressure and volume are inversely related. No units are given as Boyle used arbitrary units in his experiments.

There is always a slightly negative pressure within the thoracic cavity, which aids in keeping the airways of the lungs open. During inhalation, volume increases as a result of contraction of the diaphragm, and pressure decreases (according to Boyle's Law). This decrease of pressure in the thoracic cavity relative to the environment makes the cavity less than the atmosphere (Figure 20.16). Because of this drop in pressure, air rushes into the respiratory passages. To increase the volume of the lungs, the chest wall expands. This results from the contraction of the **intercostal muscles**, the muscles that are connected to the rib cage. Lung volume expands because the diaphragm contracts and the intercostals muscles contract, thus expanding the thoracic cavity. This increase in the volume of the thoracic cavity lowers pressure compared to the atmosphere, so air rushes into the lungs, thus increasing its volume. The resulting increase in volume is largely attributed to an increase in alveolar space, because the bronchioles and bronchi are stiff structures that do not change in size.

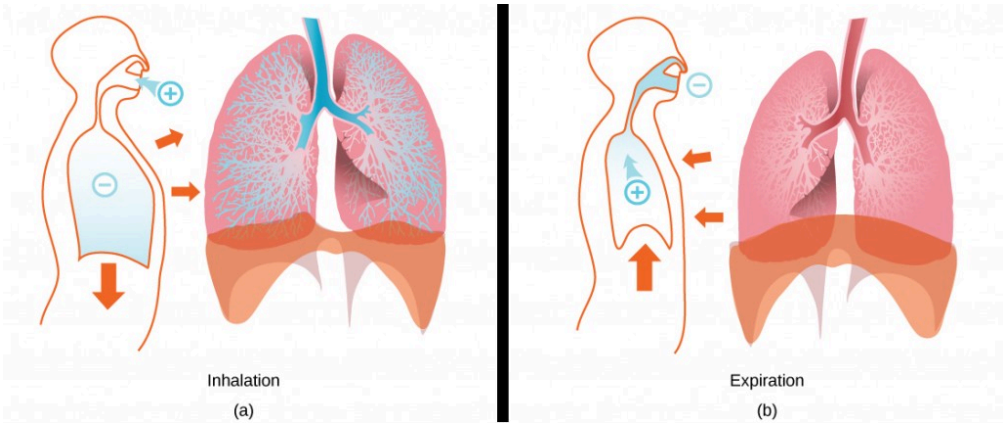


Figure 20.16. The lungs, chest wall, and diaphragm are all involved in respiration, both (a) inhalation and (b) expiration. (credit: modification of work by Mariana Ruiz Villareal)

The chest wall expands out and away from the lungs. The lungs are elastic; therefore, when air fills the lungs, the **elastic recoil** within the tissues of the lung exerts pressure back toward the interior of the lungs. These outward and inward forces compete to inflate and deflate the lung with every breath. Upon exhalation, the lungs recoil to force the air out of the lungs, and the intercostal muscles relax, returning the chest wall back to its original position (Figure 20.16 b). The diaphragm also relaxes and moves higher into the thoracic cavity. This increases the pressure within the thoracic cavity relative to the environment, and air rushes out of the lungs. The movement of air out of the lungs is a passive event. No muscles are contracting to expel the air.

Each lung is surrounded by an invaginated sac. The layer of tissue that covers the lung and dips into spaces is called the visceral **pleura**. A second layer of parietal pleura lines the interior of the thorax (Figure 20.17). The space between these layers, the **intrapleural space,** contains a small amount of fluid that protects the tissue and reduces the friction generated from rubbing the tissue layers together as the lungs contract and relax.**Pleurisy** results when these layers of tissue become inflamed; it is painful because the inflammation increases the pressure within the thoracic cavity and reduces the volume of the lung.

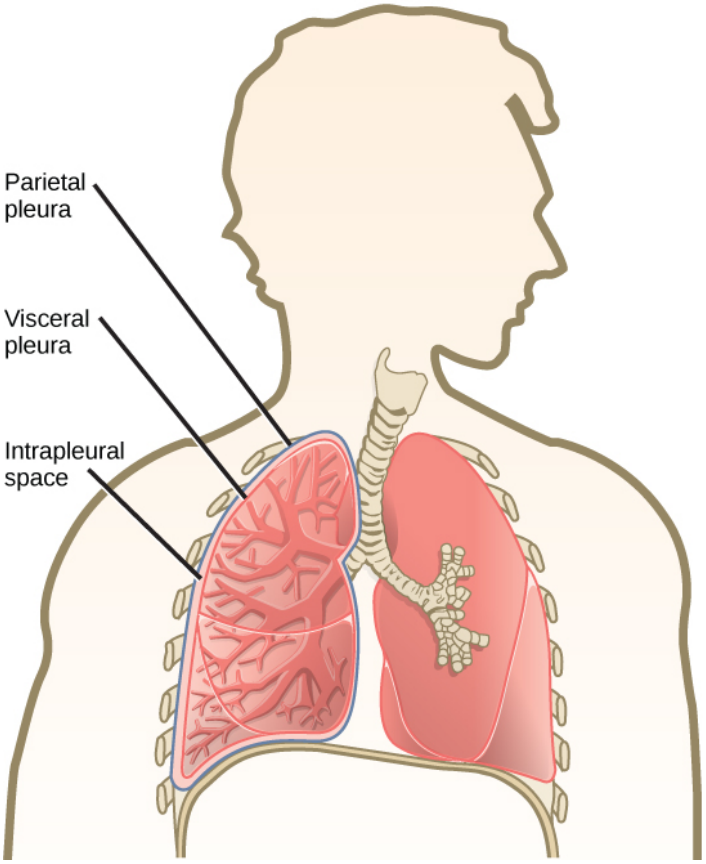


Figure 20.17. A tissue layer called pleura surrounds the lung and interior of the thoracic cavity. (credit: modification of work by NCI)

Concept in Action

View how Boyle's Law is related to breathing and watch this video on Boyle's Law.

The Work of Breathing

The number of breaths per minute is the **respiratory rate**. On average, under non-exertion conditions, the human respiratory rate is 12–15 breaths/minute. The respiratory rate contributes to the **alveolar ventilation**, or how much air moves into and out of the alveoli. Alveolar ventilation prevents carbon dioxide buildup in the alveoli. There are two ways to keep the alveolar ventilation constant: increase the respiratory rate while decreasing the tidal volume of air per breath (shallow breathing), or decrease

the respiratory rate while increasing the tidal volume per breath. In either case, the ventilation remains the same, but the work done and type of work needed are quite different. Both tidal volume and respiratory rate are closely regulated when oxygen demand increases.

There are two types of work conducted during respiration, flow-resistive and elastic work. **Flow-resistive** refers to the work of the alveoli and tissues in the lung, whereas **elastic work** refers to the work of the intercostal muscles, chest wall, and diaphragm. Increasing the respiration rate increases the flow-resistive work of the airways and decreases the elastic work of the muscles. Decreasing the respiratory rate reverses the type of work required.

Surfactant

The air-tissue/water interface of the alveoli has a high surface tension. This surface tension is similar to the surface tension of water at the liquid-air interface of a water droplet that results in the bonding of the water molecules together. **Surfactant** is a complex mixture of phospholipids and lipoproteins that works to reduce the surface tension that exists between the alveoli tissue and the air found within the alveoli. By lowering the surface tension of the alveolar fluid, it reduces the tendency of alveoli to collapse.

Surfactant works like a detergent to reduce the surface tension and allows for easier inflation of the airways. When a balloon is first inflated, it takes a large amount of effort to stretch the plastic and start to inflate the balloon. If a little bit of detergent was applied to the interior of the balloon, then the amount of effort or work needed to begin to inflate the balloon would decrease, and it would become much easier to start blowing up the balloon. This same principle applies to the airways. A small amount of surfactant to the airway tissues reduces the effort or work needed to inflate those airways. Babies born prematurely sometimes do not produce enough surfactant. As a result, they suffer from **respiratory distress syndrome**, because it requires more effort to inflate their lungs. Surfactant is also important for preventing collapse of small alveoli relative to large alveoli.

Lung Resistance and Compliance

Pulmonary diseases reduce the rate of gas exchange into and out of the lungs. Two main causes of decreased gas exchange are **compliance** (how elastic the lung is) and **resistance**(how much obstruction exists in the airways). A change in either can dramatically alter breathing and the ability to take in oxygen and release carbon dioxide.

Examples of **restrictive diseases** are respiratory distress syndrome and pulmonary fibrosis. In both diseases, the airways are less compliant and they are stiff or fibrotic. There is a decrease in compliance because the lung tissue cannot bend and move. In these types of restrictive diseases, the intrapleural pressure is more positive and the airways collapse upon exhalation, which traps air in the lungs. Forced or **functional vital capacity (FVC)**, which is the amount of air that can be forcibly exhaled after taking the deepest breath possible, is much lower than in normal patients, and the time it takes to exhale most of the air is greatly prolonged (Figure 20.18). A patient suffering from these diseases cannot exhale the normal amount of air.

Obstructive diseases and conditions include emphysema, asthma, and pulmonary edema. In

emphysema, which mostly arises from smoking tobacco, the walls of the alveoli are destroyed, decreasing the surface area for gas exchange. The overall compliance of the lungs is increased, because as the alveolar walls are damaged, lung elastic recoil decreases due to a loss of elastic fibers, and more air is trapped in the lungs at the end of exhalation. Asthma is a disease in which inflammation is triggered by environmental factors. Inflammation obstructs the airways. The obstruction may be due to edema (fluid accumulation), smooth muscle spasms in the walls of the bronchioles, increased mucus secretion, damage to the epithelia of the airways, or a combination of these events. Those with asthma or edema experience increased occlusion from increased inflammation of the airways. This tends to block the airways, preventing the proper movement of gases (Figure 20.18). Those with obstructive diseases have large volumes of air trapped after exhalation and breathe at a very high lung volume to compensate for the lack of airway recruitment.

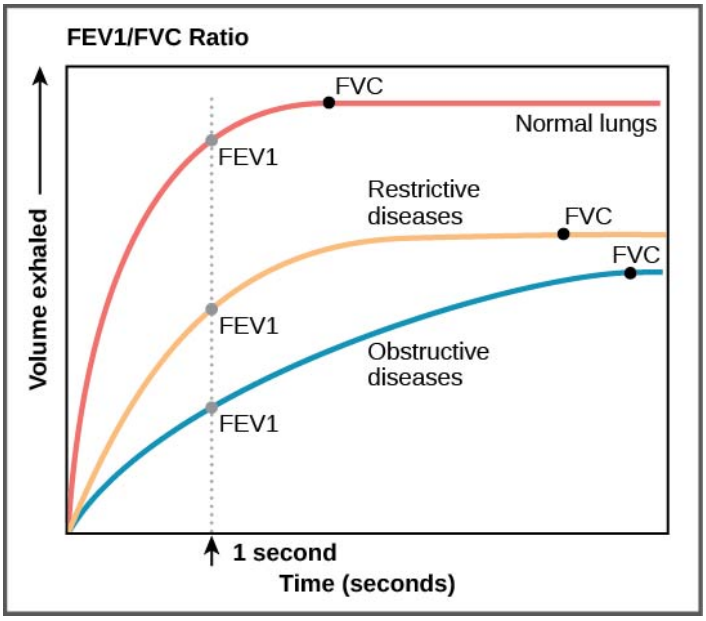


Figure 20.18.
The ratio of FEV1 (the amount of air that can be forcibly exhaled in one second after taking a deep breath) to FVC (the total amount of air that can be forcibly exhaled) can be used to diagnose whether a person has restrictive or obstructive lung disease. In restrictive lung disease, FVC is reduced but airways are not obstructed, so the person is able to expel air reasonably fast. In obstructive lung disease, airway obstruction results in slow exhalation as well as reduced FVC. Thus, the FEV1/FVC ratio is lower in persons with obstructive lung disease (less than 69 percent) than in persons with restrictive disease (88 to 90 percent).

Dead Space: V/Q Mismatch

Pulmonary circulation pressure is very low compared to that of the systemic circulation. It is also independent of cardiac output. This is because of a phenomenon called **recruitment**, which is the process of opening airways that normally remain closed when cardiac output increases. As cardiac output increases, the number of capillaries and arteries that are perfused (filled with blood) increases. These capillaries and arteries are not always in use but are ready if needed. At times, however, there is

a mismatch between the amount of air (ventilation, V) and the amount of blood (perfusion, Q) in the lungs. This is referred to as **ventilation/perfusion (V/Q) mismatch**.

There are two types of V/Q mismatch. Both produce **dead space**, regions of broken down or blocked lung tissue. Dead spaces can severely impact breathing, because they reduce the surface area available for gas diffusion. As a result, the amount of oxygen in the blood decreases, whereas the carbon dioxide level increases. Dead space is created when no ventilation and/or perfusion takes place. **Anatomical dead space** or anatomical shunt, arises from an anatomical failure, while **physiological dead space**or physiological shunt, arises from a functional impairment of the lung or arteries.

An example of an anatomical shunt is the effect of gravity on the lungs. The lung is particularly susceptible to changes in the magnitude and direction of gravitational forces. When someone is standing or sitting upright, the pleural pressure gradient leads to increased ventilation further down in the lung. As a result, the intrapleural pressure is more negative at the base of the lung than at the top, and more air fills the bottom of the lung than the top. Likewise, it takes less energy to pump blood to the bottom of the lung than to the top when in a prone position. Perfusion of the lung is not uniform while standing or sitting. This is a result of hydrostatic forces combined with the effect of airway pressure. An anatomical shunt develops because the ventilation of the airways does not match the perfusion of the arteries surrounding those airways. As a result, the rate of gas exchange is reduced. Note that this does not occur when lying down, because in this position, gravity does not preferentially pull the bottom of the lung down.

A physiological shunt can develop if there is infection or edema in the lung that obstructs an area. This will decrease ventilation but not affect perfusion; therefore, the V/Q ratio changes and gas exchange is affected.

The lung can compensate for these mismatches in ventilation and perfusion. If ventilation is greater than perfusion, the arterioles dilate and the bronchioles constrict. This increases perfusion and reduces ventilation. Likewise, if ventilation is less than perfusion, the arterioles constrict and the bronchioles dilate to correct the imbalance.

Concept in Action

>Visit this site to view the mechanics of breathing.

Summary

The structure of the lungs and thoracic cavity control the mechanics of breathing. Upon inspiration, the diaphragm contracts and lowers. The intercostal muscles contract and expand the chest wall outward.

The intrapleural pressure drops, the lungs expand, and air is drawn into the airways. When exhaling, the intercostal muscles and diaphragm relax, returning the intrapleural pressure back to the resting state. The lungs recoil and airways close. The air passively exits the lung. There is high surface tension at the air-airway interface in the lung. Surfactant, a mixture of phospholipids and lipoproteins, acts like a detergent in the airways to reduce surface tension and allow for opening of the alveoli.

Breathing and gas exchange are both altered by changes in the compliance and resistance of the lung. If the compliance of the lung decreases, as occurs in restrictive diseases like fibrosis, the airways stiffen and collapse upon exhalation. Air becomes trapped in the lungs, making breathing more difficult. If resistance increases, as happens with asthma or emphysema, the airways become obstructed, trapping air in the lungs and causing breathing to become difficult. Alterations in the ventilation of the airways or perfusion of the arteries can affect gas exchange. These changes in ventilation and perfusion, called V/Q mismatch, can arise from anatomical or physiological changes.

Exercises

1. How would paralysis of the diaphragm alter inspiration?
    1. It would prevent contraction of the intercostal muscles.
    2. It would prevent inhalation because the intrapleural pressure would not change.
    3. It would decrease the intrapleural pressure and allow more air to enter the lungs.
    4. It would slow expiration because the lung would not relax.
2. Restrictive airway diseases _____.
    1. increase the compliance of the lung
    2. decrease the compliance of the lung
    3. increase the lung volume
    4. decrease the work of breathing
3. Alveolar ventilation remains constant when _____.
    1. the respiratory rate is increased while the volume of air per breath is decreased
    2. the respiratory rate and the volume of air per breath are increased
    3. the respiratory rate is decreased while increasing the volume per breath
    4. both a and c
4. How would increased airway resistance affect intrapleural pressure during inhalation?
5. Explain how a puncture to the thoracic cavity (from a knife wound, for instance) could alter the ability to inhale.
6. When someone is standing, gravity stretches the bottom of the lung down toward the floor to a greater extent than the top of the lung. What implication could this have on the flow of air in the lungs? Where does gas exchange occur in the lungs?

Answers

1. B
2. B
3. D
4. Increased airway resistance increases the volume and pressure in the lung; therefore, the intrapleural pressure would be less negative and breathing would be more difficult.
5. A puncture to the thoracic cavity would equalize the pressure inside the thoracic cavity to the outside environment. For the lung to function properly, the intrapleural pressure must be negative. This is caused by the contraction of the diaphragm pulling the lungs down and drawing air into the lungs.
6. The lung is particularly susceptible to changes in the magnitude and direction of gravitational forces. When someone is standing or sitting upright, the pleural pressure gradient leads to increased ventilation further down in the lung.

Glossary

alveolar ventilation
how much air is in the alveoli

anatomical dead space
(also, anatomical shunt) region of the lung that lacks proper ventilation/perfusion due to an anatomical block

compliance
measurement of the elasticity of the lung

dead space
area in the lung that lacks proper ventilation or perfusion

elastic recoil
property of the lung that drives the lung tissue inward

elastic work
work conducted by the intercostal muscles, chest wall, and diaphragm

flow-resistive
work of breathing performed by the alveoli and tissues in the lung

functional vital capacity (FVC)
amount of air that can be forcibly exhaled after taking the deepest breath possible

intercostal muscle
muscle connected to the rib cage that contracts upon inspiration

physiological dead space
(also, physiological shunt) region of the lung that lacks proper ventilation/perfusion due to a physiological change in the lung (like inflammation or edema)

pleura
tissue layer that surrounds the lungs and lines the interior of the thoracic cavity

pleurisy
painful inflammation of the pleural tissue layers

recruitment
process of opening airways that normally remain closed when the cardiac output increases

respiratory rate
number of breaths per minute

ventilation/perfusion (V/Q) mismatch
region of the lung that lacks proper alveolar ventilation (V) and/or arterial perfusion (Q)

20.4 Transport of Gases in Human Bodily Fluids

Learning Objectives

By the end of this section, you will be able to:

- Describe how oxygen is bound to hemoglobin and transported to body tissues
- Explain how carbon dioxide is transported from body tissues to the lungs

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 20.19). 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.

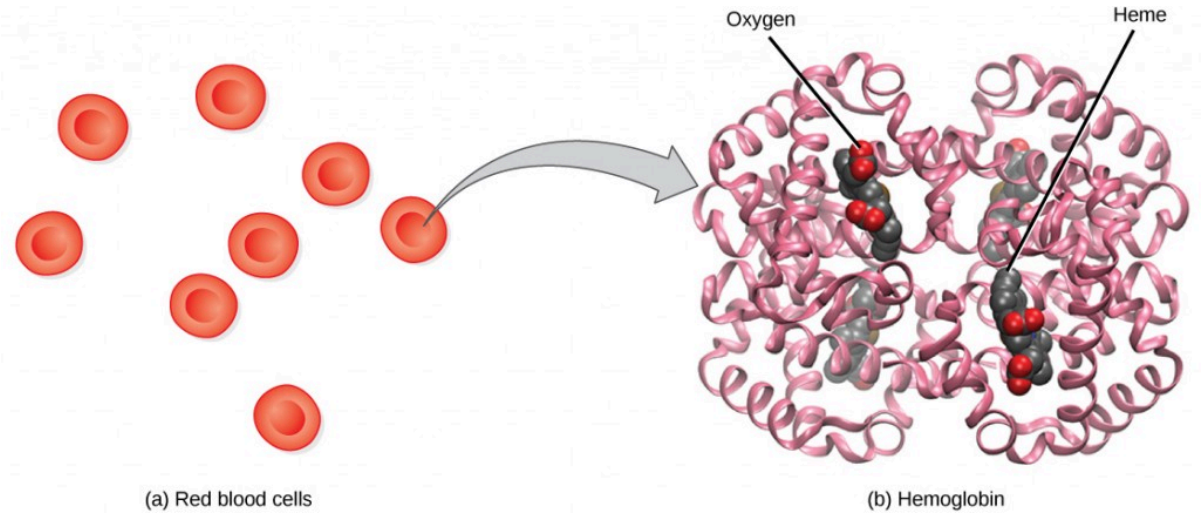


Figure 20.19. The protein inside (a) red blood cells that carries oxygen to cells and carbon dioxide to the lungs is (b) hemoglobin. Hemoglobin is made up of four symmetrical subunits and four heme groups. Iron associated with the heme binds oxygen. It is the iron in hemoglobin that gives blood its red color.

It is easier to bind a second and third oxygen molecule to Hb than the first molecule. This is because the hemoglobin molecule changes its shape, or conformation, as oxygen binds. The fourth oxygen is then more difficult to bind. The binding of oxygen to hemoglobin can be plotted as a function of the partial pressure of oxygen in the blood (x-axis) versus the relative Hb-oxygen saturation (y-axis). The resulting graph—an **oxygen dissociation curve**—is sigmoidal, or S-shaped (Figure 20.20). As the partial pressure of oxygen increases, the hemoglobin becomes increasingly saturated with oxygen.

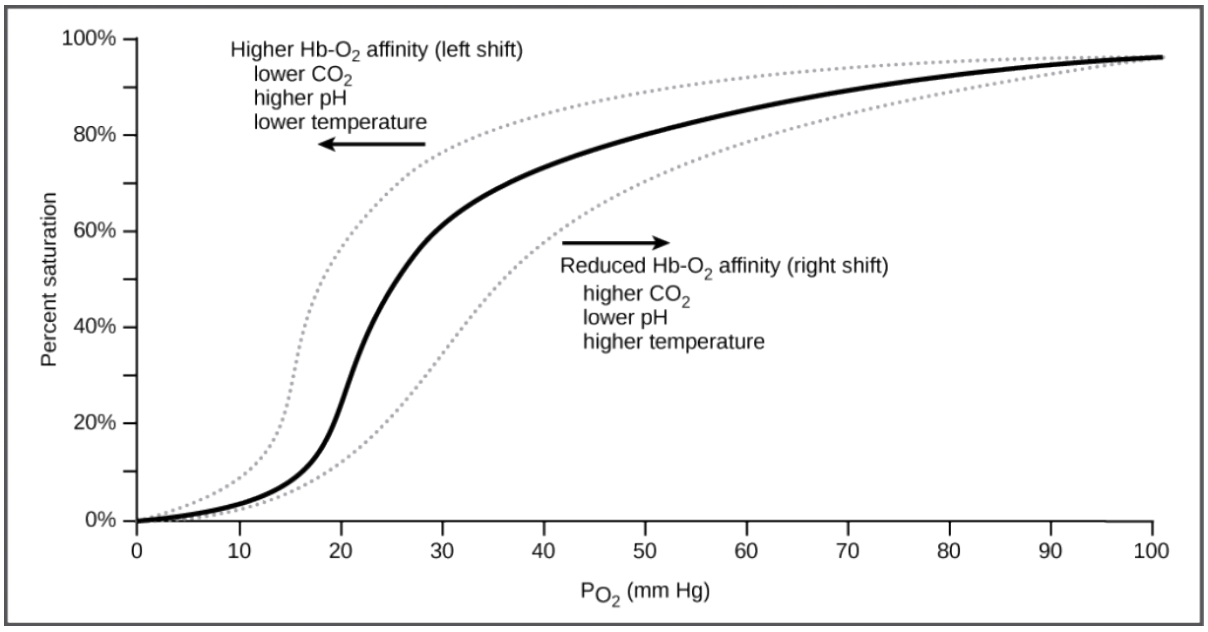


Figure 20.20. The oxygen dissociation curve demonstrates that, as the partial pressure of oxygen increases, more oxygen binds hemoglobin. However, the affinity of hemoglobin for oxygen may shift to the left or the right depending on environmental conditions.

The kidneys are responsible for removing excess H+ ions from the blood. If the kidneys fail, what would happen to blood pH and to hemoglobin affinity for oxygen?

Factors That Affect Oxygen Binding

The **oxygen-carrying capacity** of hemoglobin determines how much oxygen is carried in the blood. In addition to P_{O2}, other environmental factors and diseases can affect oxygen carrying capacity and delivery.

Carbon dioxide levels, blood pH, and body temperature affect oxygen-carrying capacity (Figure 20.20). When carbon dioxide is in the blood, it reacts with water to form bicarbonate (HCO^-_3) and hydrogen ions (H^+). As the level of carbon dioxide in the blood increases, more H^+ is produced and the pH decreases. This increase in carbon dioxide and subsequent decrease in pH reduce the affinity of hemoglobin for oxygen. The oxygen dissociates from the Hb molecule, shifting the oxygen dissociation curve to the right. Therefore, more oxygen is needed to reach the same hemoglobin saturation level as when the pH was higher. A similar shift in the curve also results from an increase in body temperature. Increased temperature, such as from increased activity of skeletal muscle, causes the affinity of hemoglobin for oxygen to be reduced.

Diseases like sickle cell anemia and thalassemia decrease the blood's ability to deliver oxygen to tissues and its oxygen-carrying capacity. In **sickle cell anemia**, the shape of the red blood cell is crescent-shaped, elongated, and stiffened, reducing its ability to deliver oxygen (Figure 20.21). In this form, red blood cells cannot pass through the capillaries. This is painful when it occurs. **Thalassemia** is a rare genetic disease caused by a defect in either the alpha or the beta subunit of Hb. Patients with thalassemia produce a high number of red blood cells, but these cells have lower-than-normal levels of hemoglobin. Therefore, the oxygen-carrying capacity is diminished.

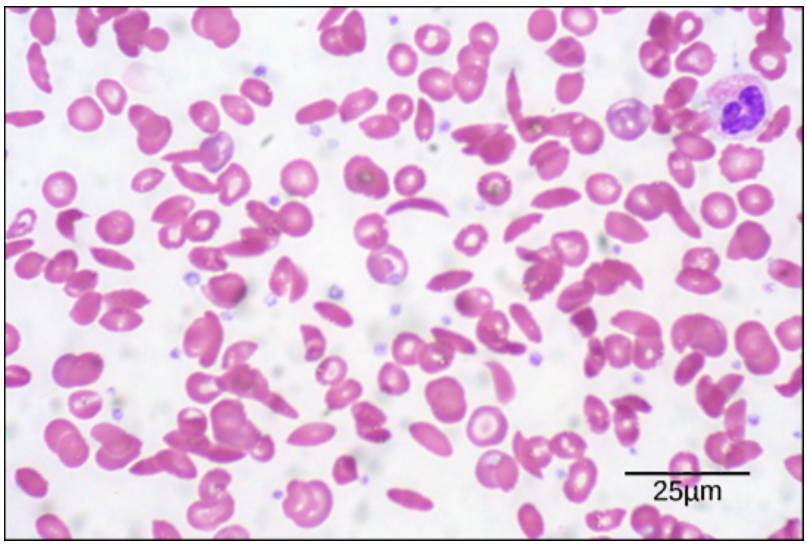


Figure 20.21.
Individuals with sickle cell anemia have crescent-shaped red blood cells. (credit: modification of work by Ed Uthman; scale-bar data from Matt Russell)

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 (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 liquid component of the blood in exchange for a chloride ion (Cl^-); this is called the **bicarbonate (HCO−3) ion**. 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.

Carbon Monoxide Poisoning

While carbon dioxide can readily associate and dissociate from hemoglobin, other molecules such as carbon monoxide (CO) cannot. Carbon monoxide has a greater affinity for hemoglobin than oxygen. Therefore, when carbon monoxide is present, it binds to hemoglobin preferentially over oxygen. As a result, oxygen cannot bind to hemoglobin, so very little oxygen is transported through the body (Figure 20.22). Carbon monoxide is a colorless, odorless gas and is therefore difficult to detect. It is produced by gas-powered vehicles and tools. Carbon monoxide can cause headaches, confusion, and nausea; long-term exposure can cause brain damage or death. Administering 100 percent (pure) oxygen is the usual treatment for carbon monoxide poisoning. Administration of pure oxygen speeds up the separation of carbon monoxide from hemoglobin.

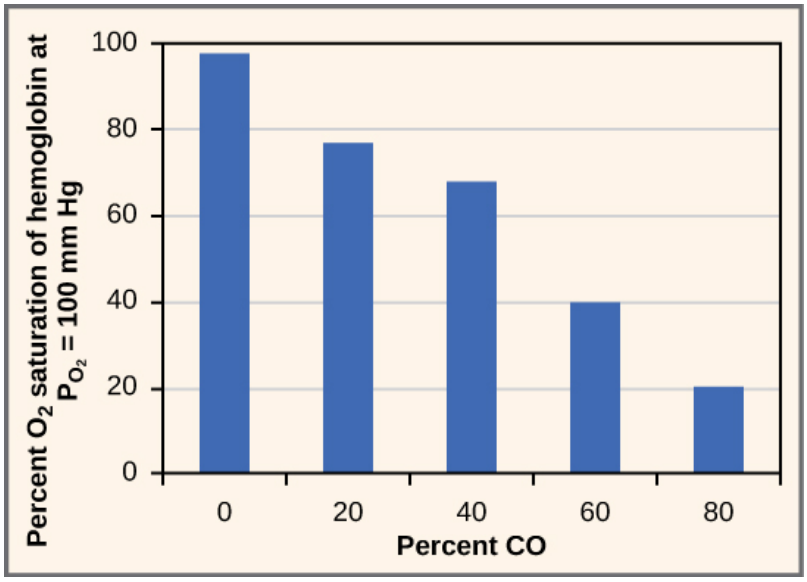


Figure 20.22. As percent CO increases, the oxygen saturation of hemoglobin decreases.

Summary

Hemoglobin is a protein found in red blood cells that is comprised of two alpha and two beta subunits that surround an iron-containing heme group. Oxygen readily binds this heme group. The ability of oxygen to bind increases as more oxygen molecules are bound to heme. Disease states and altered conditions in the body can affect the binding ability of oxygen, and increase or decrease its ability to dissociate from hemoglobin.

Carbon dioxide can be transported through the blood via three methods. It is dissolved directly in the blood, bound to plasma proteins or hemoglobin, or converted into bicarbonate. The majority of carbon dioxide is transported as part of the bicarbonate system. Carbon dioxide diffuses into red blood cells. Inside, carbonic anhydrase converts carbon dioxide into carbonic acid (H_2CO_3), which is subsequently hydrolyzed into bicarbonate (HCO^-_3) and H^+. The H^+ ion binds to hemoglobin in red blood cells, and bicarbonate is transported out of the red blood cells in exchange for a chloride ion. This is called the chloride shift. Bicarbonate leaves the red blood cells and enters the blood plasma. In the lungs, bicarbonate is transported back into the red blood cells in exchange for chloride. The H^+ dissociates from hemoglobin and combines with bicarbonate to form carbonic acid with the help of carbonic anhydrase, which further catalyzes the reaction to convert carbonic acid back into carbon dioxide and water. The carbon dioxide is then expelled from the lungs.

Exercises

1. The kidneys are responsible for removing excess H+ ions from the blood. If the kidneys fail, what would happen to blood pH and to hemoglobin affinity for oxygen?
2. Which of the following will NOT facilitate the transfer of oxygen to tissues?

1. decreased body temperature
2. decreased pH of the blood
3. increased carbon dioxide
4. increased exercise

3. The majority of carbon dioxide in the blood is transported by _____.

    1. binding to hemoglobin
    2. dissolution in the blood
    3. conversion to bicarbonate
    4. binding to plasma proteins

4. The majority of oxygen in the blood is transported by _____.

    1. dissolution in the blood
    2. being carried as bicarbonate ions
    3. binding to blood plasma
    4. binding to hemoglobin

5. What would happen if no carbonic anhydrase were present in red blood cells?
6. How does the administration of 100 percent oxygen save a patient from carbon monoxide poisoning? Why wouldn't giving carbon dioxide work?

Answers

1. The blood pH will drop and hemoglobin affinity for oxygen will decrease.
2. A
3. C
4. D
5. Without carbonic anhydrase, carbon dioxide would not be hydrolyzed into carbonic acid or bicarbonate. Therefore, very little carbon dioxide (only 15 percent) would be transported in the blood away from the tissues.
6. Carbon monoxide has a higher affinity for hemoglobin than oxygen. This means that carbon monoxide will preferentially bind to hemoglobin over oxygen. Administration of 100 percent oxygen is an effective therapy because at that concentration, oxygen will displace the carbon monoxide from the hemoglobin.

Glossary

bicarbonate (HCO^-_3) ion
ion created when carbonic acid dissociates into H^+ and (HCO^-_3)

bicarbonate buffer system
system in the blood that absorbs carbon dioxide and regulates pH levels

carbaminohemoglobin
molecule that forms when carbon dioxide binds to hemoglobin

carbonic anhydrase (CA)
enzyme that catalyzes carbon dioxide and water into carbonic acid

heme group
centralized iron-containing group that is surrounded by the alpha and beta subunits of hemoglobin

hemoglobin
molecule in red blood cells that can bind oxygen, carbon dioxide, and carbon monoxide

oxygen dissociation curve
curve depicting the affinity of oxygen for hemoglobin

oxygen-carrying capacity
amount of oxygen that can be transported in the blood

sickle cell anemia
genetic disorder that affects the shape of red blood cells, and their ability to transport oxygen and move through capillaries

thalassemia
rare genetic disorder that results in mutation of the alpha or beta subunits of hemoglobin, creating smaller red blood cells with less hemoglobin

Chapter 21. The Circulatory System

Figure 21.1. Just as highway systems transport people and goods through a complex network, the circulatory system transports nutrients, gases, and wastes throughout the animal body. (credit: modification of work by Andrey Belenko)

Introduction

Most animals are complex multicellular organisms that require a mechanism for transporting nutrients throughout their bodies and removing waste products. The circulatory system has evolved over time from simple diffusion through cells in the early evolution of animals to a complex network of blood vessels that reach all parts of the human body. This extensive network supplies the cells, tissues, and organs with oxygen and nutrients, and removes carbon dioxide and waste, which are byproducts of respiration.

At the core of the human circulatory system is the heart. The size of a clenched fist, the human heart is protected beneath the rib cage. Made of specialized and unique cardiac muscle, it pumps blood throughout the body and to the heart itself. Heart contractions are driven by intrinsic electrical impulses that the brain and endocrine hormones help to regulate. Understanding the heart’s basic anatomy and function is important to understanding the body’s circulatory and respiratory systems.

Gas exchange is one essential function of the circulatory system. A circulatory system is not needed in organisms with no specialized respiratory organs because oxygen and carbon dioxide diffuse directly between their body tissues and the external environment. However, in organisms that possess lungs and gills, oxygen must be transported from these specialized respiratory organs to the body tissues via a circulatory system. Therefore, circulatory systems have had to evolve to accommodate the great diversity of body sizes and body types present among animals.

21.1. Overview of the Circulatory System

Learning Objectives

By the end of this section, you will be able to:

- Describe an open and closed circulatory system
- Describe interstitial fluid and hemolymph
- Compare and contrast the organization and evolution of the vertebrate circulatory system.

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-loop 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 21.2**a**. As opposed to a closed system, arthropods—including insects, crustaceans, and most mollusks—have an open circulatory system, as illustrated in Figure 21.2**b**. In an **open circulatory system**, the blood is not enclosed in the blood vessels but is pumped into a cavity called a **hemocoel** and 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 metabolically active organs and tissues that require high levels of oxygen. In fact, one reason that insects with wing spans of up to two feet wide (70 cm) are not around today is probably because they were outcompeted by the arrival of birds 150 million years ago. Birds, having a closed circulatory system, are thought to have moved more agilely, allowing them to get food faster and possibly to prey on the insects.

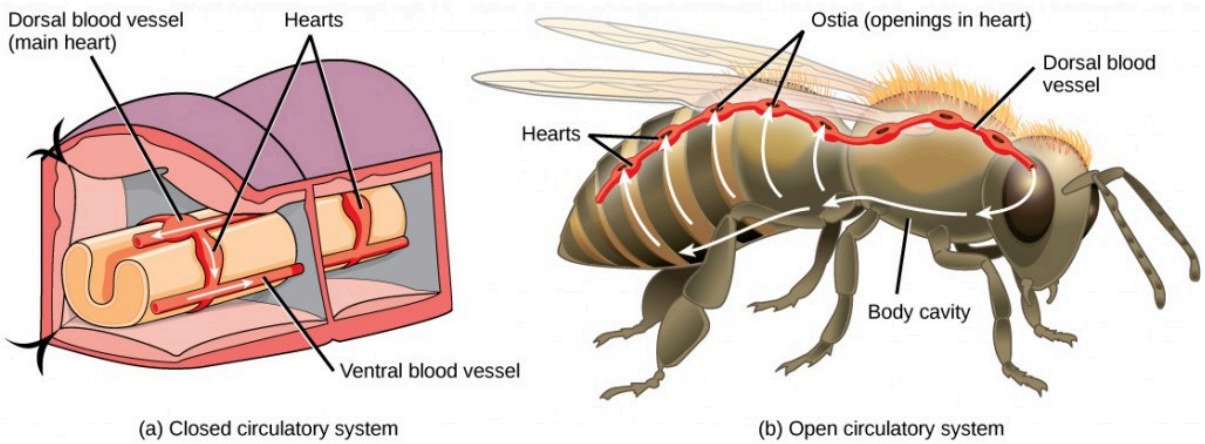


Figure 21.2. In (a) closed circulatory systems, the heart pumps blood through vessels that are separate from the interstitial fluid of the body. Most vertebrates and some invertebrates, like this annelid earthworm, have a closed circulatory system. In (b) open circulatory systems, a fluid called hemolymph 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 21.3**a**. 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 21.3**b**. Exchange of fluids is assisted by the pulsing of the jellyfish body.

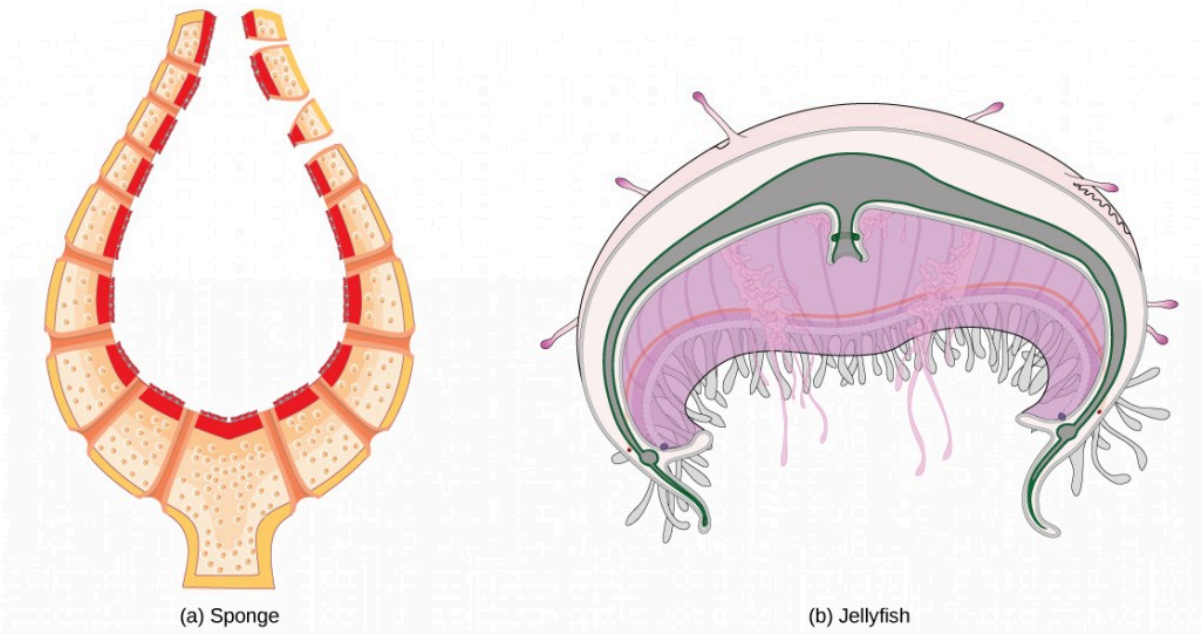


Figure 21.3. Simple animals consisting of a single cell layer such as the (a) sponge or only a few cell layers such as the (b) jellyfish do not have a circulatory system. Instead, gases, nutrients, and wastes are exchanged by diffusion.

For more complex organisms, diffusion is not efficient for cycling gases, nutrients, and waste effectively through the body; therefore, more complex circulatory systems evolved. 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 a characteristic of 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 21.4 illustrates the basic circulatory systems of some vertebrates: fish, amphibians, reptiles, and mammals.

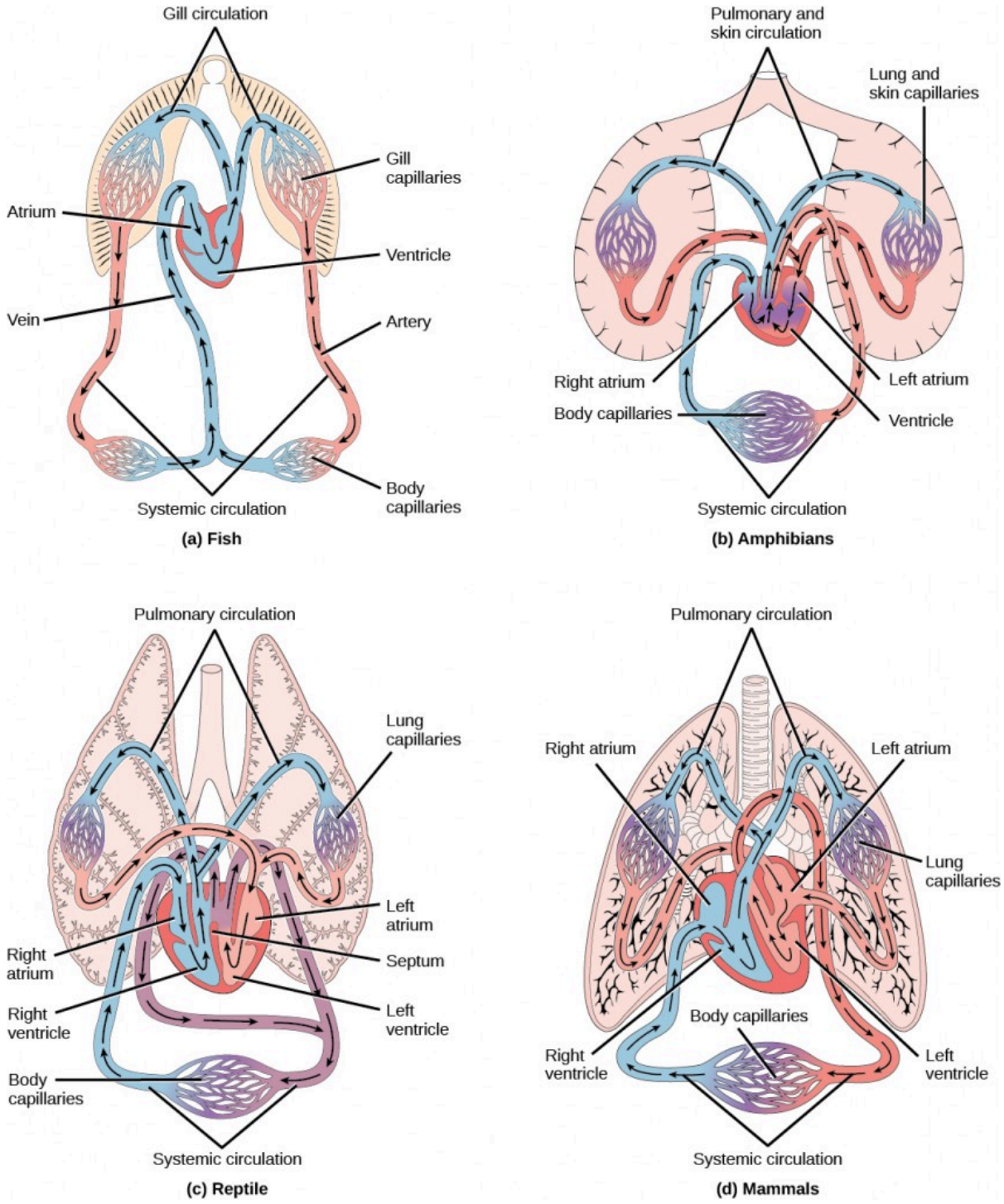


Figure 21.4. (a) Fish have the simplest circulatory systems of the vertebrates: blood flows unidirectionally from the two-chambered heart through the gills and then the rest of the body. (b) Amphibians have two circulatory routes: one for oxygenation of the blood through the lungs and skin, and the other 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 21.4**a** 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 **gill circulation**. The blood then continues through the rest of the body before arriving back at the atrium; this is called **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 **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 21.4**b**, 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 efficiency of oxygenation. 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 21.4**c**. The ventricle is divided more effectively by a partial septum, which results in 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 evolutionarily successful 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 21.4**d**. The oxygenated blood is 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 a three-chambered heart. The independent evolution of the same or a similar biological trait is referred to as convergent evolution.

Summary

In most animals, the circulatory system is used to transport blood through the body. Some primitive animals use diffusion for the exchange of water, nutrients, and gases. However, complex organisms use the circulatory system to carry gases, nutrients, and waste through the body. Circulatory systems may be open (mixed with the interstitial fluid) or closed (separated from the interstitial fluid). Closed circulatory systems are a characteristic of vertebrates; however, there are significant differences in the structure of the heart and the circulation of blood between the different vertebrate groups due to adaptions during evolution and associated differences in anatomy. Fish have a two-chambered heart with unidirectional circulation. Amphibians have a three-chambered heart, which has some mixing of the blood, and they have double circulation. Most non-avian reptiles have a three-chambered heart, but have little mixing of the blood; they have double circulation. Mammals and birds have a four-chambered heart with no mixing of the blood and double circulation.

Exercises

1. Which of the following statements about the circulatory system is false?
    1. Blood in the pulmonary vein is deoxygenated.
    2. Blood in the inferior vena cava is deoxygenated.
    3. Blood in the pulmonary artery is deoxygenated.
    4. Blood in the aorta is oxygenated.
2. Which of the following statements about the heart is false?
    1. The mitral valve separates the left ventricle from the left atrium.
    2. Blood travels through the bicuspid valve to the left atrium.
    3. Both the aortic and the pulmonary valves are semilunar valves.
    4. The mitral valve is an atrioventricular valve.
3. Varicose veins are veins that become enlarged because the valves no longer close properly, allowing blood to flow backward. Varicose veins are often most prominent on the legs. Why do you think this is the case?
4. Why are open circulatory systems advantageous to some animals?
    1. They use less metabolic energy.
    2. They help the animal move faster.
    3. They do not need a heart.
    4. They help large insects develop.
5. Some animals use diffusion instead of a circulatory system. Examples include:
    1. birds and jellyfish
    2. flatworms and arthropods

3. mollusks and jellyfish
4. None of the above

6. Blood flow that is directed through the lungs and back to the heart is called _____.
    1. unidirectional circulation
    2. gill circulation
    3. pulmonary circulation
    4. pulmocutaneous circulation
7. Describe a closed circulatory system.
8. Describe systemic circulation.

Answers

1. C
2. B
3. Blood in the legs is farthest away from the heart and has to flow up to reach it.
4. A
5. D
6. C
7. A closed circulatory system is a closed-loop system, in which blood is not free in a cavity. Blood is separate from the bodily interstitial fluid and contained within blood vessels. In this type of system, blood circulates unidirectionally from the heart around the systemic circulatory route, and then returns to the heart.
8. Systemic circulation flows through the systems of the body. The blood flows away from the heart to the brain, liver, kidneys, stomach, and other organs, the limbs, and the muscles of the body; it then returns to the heart.

Glossary

atrium
(plural: atria) chamber of the heart that receives blood from the veins and sends blood to the ventricles

closed circulatory system
system in which the blood is separated from the bodily interstitial fluid and contained in blood vessels

double circulation
flow of blood in two circuits: the pulmonary circuit through the lungs and the systemic circuit through the organs and body

gill circulation
circulatory system that is specific to animals with gills for gas exchange; the blood flows through the

gills for oxygenation

hemocoel
cavity into which blood is pumped in an open circulatory system

hemolymph
mixture of blood and interstitial fluid that is found in insects and other arthropods as well as most mollusks

interstitial fluid
fluid between cells

ostium
(plural: ostia) holes between blood vessels that allow the movement of hemolymph through the body of insects, arthropods, and mollusks with open circulatory systems

pulmonary circulation
flow of blood away from the heart through the lungs where oxygenation occurs and then returns to the heart again

systemic circulation
flow of blood away from the heart to the brain, liver, kidneys, stomach, and other organs, the limbs, and the muscles of the body, and then the return of this blood to the heart

unidirectional circulation
flow of blood in a single circuit; occurs in fish where the blood flows through the gills, then past the organs and the rest of the body, before returning to the heart

ventricle
(heart) large inferior chamber of the heart that pumps blood into arteries

21.2. Components of the Blood

Learning Objectives

By the end of this section, you will be able to:

- List the basic components of the blood
- Compare red and white blood cells
- Describe blood plasma and serum

Hemoglobin is responsible for distributing oxygen, and to a lesser extent, carbon dioxide, throughout the circulatory systems of humans, vertebrates, and many invertebrates. The blood is more than the 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**. Blood plasma is actually the dominant 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. Interstitial fluid that surrounds cells is separate from the blood, but in hemolymph, they are combined. In humans, cellular components make up approximately 45 percent of the blood and the liquid plasma 55 percent. Blood is 20 percent of a person's extracellular fluid and eight percent of weight.

The Role of Blood in the Body

Blood, like the human blood illustrated in

Figure 21.5 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.

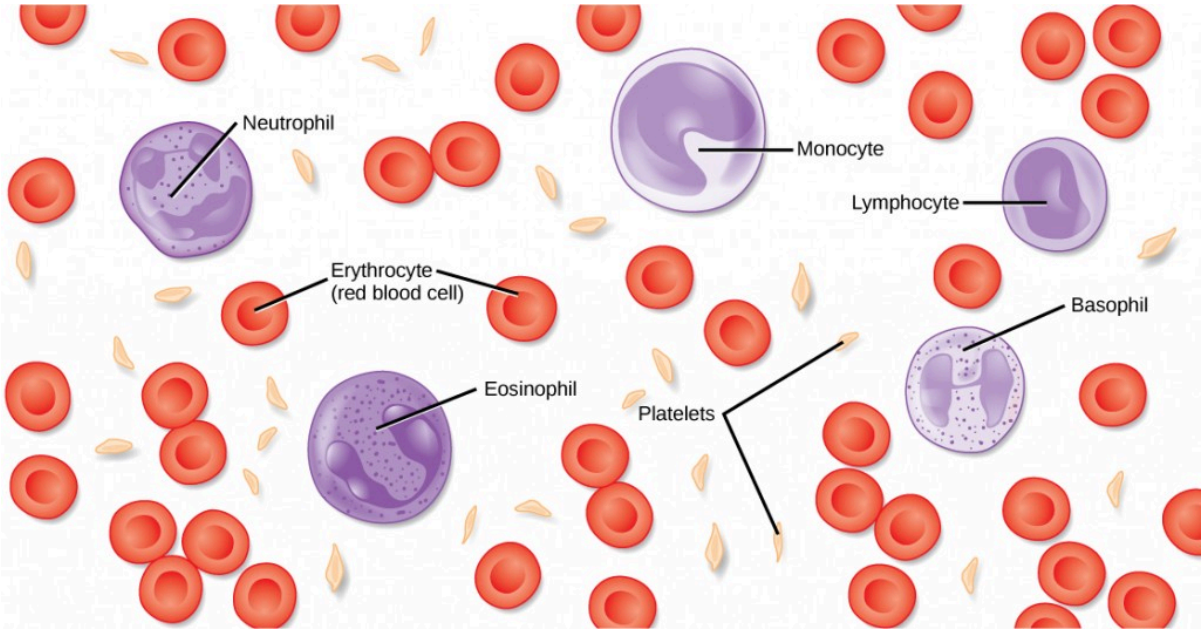


Figure 21.5. The cells and cellular components of human blood are shown. Red blood cells deliver oxygen to the cells and remove carbon dioxide. White blood cells—including neutrophils, monocytes, lymphocytes, eosinophils, and basophils—are involved in the immune response. Platelets form clots that prevent 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 formed from 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 non-avian reptiles, a nucleus is still maintained in red blood cells.

The red coloring of blood comes from the iron-containing protein hemoglobin, illustrated in Figure 21.6**a**. The principal job of this protein is to carry oxygen, but it also transports carbon dioxide as well. Hemoglobin is packed into 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. Only mammals have anucleated red blood cells, and some mammals (camels, for instance) even have nucleated red blood cells. The advantage of nucleated red blood cells is that these cells can undergo mitosis. Anucleated red blood cells metabolize anaerobically (without oxygen), making use of a primitive metabolic pathway to produce ATP and increase the efficiency of oxygen transport.

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 21.6**b** 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 21.6**c**. Despite the name, hemerythrin does not contain a heme group and its oxygen-carrying capacity is poor compared to hemoglobin.

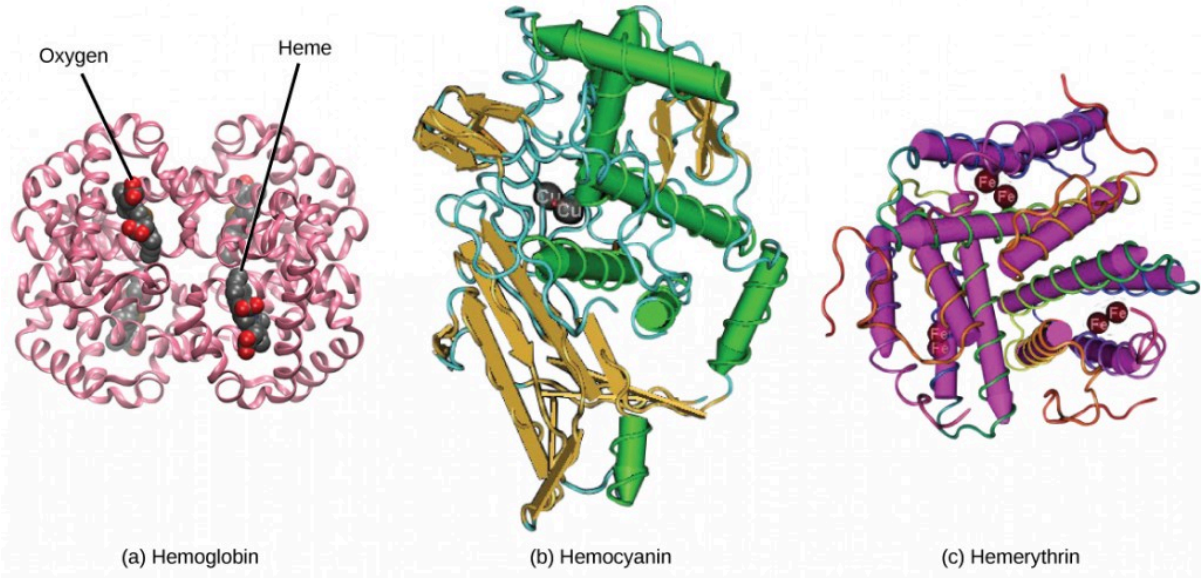


Figure 21.6. In most vertebrates, (a) hemoglobin delivers oxygen to the body and removes some carbon dioxide. Hemoglobin is composed of four protein subunits, two alpha chains and two beta chains, and a heme group that has iron associated with it. The iron reversibly associates with oxygen, and in so doing is oxidized from Fe2+ to Fe3+. 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. Studies have found that hemoglobin also binds nitrous oxide (NO). NO is a vasodilator that relaxes the blood vessels and capillaries and may help with gas exchange and the passage of red blood cells through narrow vessels. Nitroglycerin, a heart medication for angina and heart attacks, is converted to NO to help relax the blood vessels and increase oxygen flow through the body.

A characteristic of red blood cells is their glycolipid and glycoprotein coating; these are lipids and proteins that have carbohydrate molecules attached. In humans, the surface glycoproteins and glycolipids on red blood cells vary between individuals, producing the different blood types, such as A, B, and O. Red blood cells have an average life span of 120 days, at which time they are broken down and recycled in the liver and spleen by phagocytic macrophages, a type of white blood cell.

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: 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 21.7 are the granulocytes, which include the neutrophils, eosinophils, and basophils, and the agranulocytes, which include the monocytes and lymphocytes.

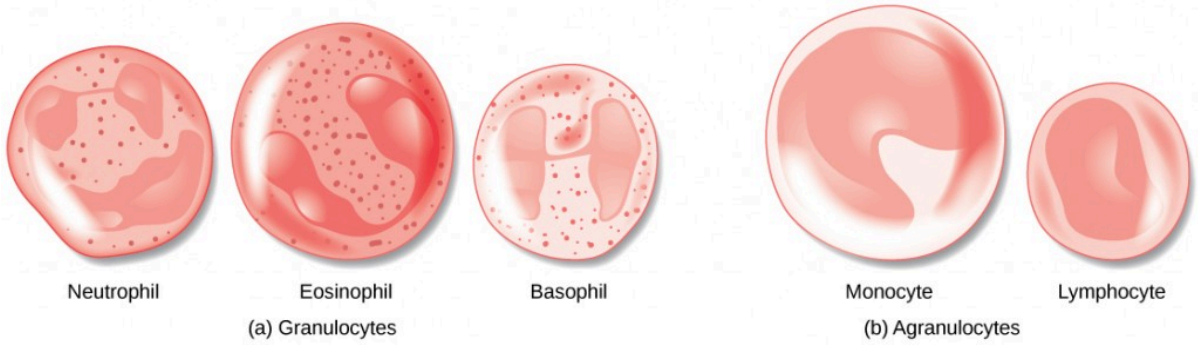


Figure 21.7. (a) Granulocytes—including neutrophils, eosinophils and basophils—are characterized by a lobed nucleus and granular inclusions in the cytoplasm. Granulocytes are typically first-responders during injury or infection. (b) Agranulocytes include lymphocytes and monocytes. Lymphocytes, including B and T cells, are responsible for adaptive immune response. Monocytes differentiate into macrophages and dendritic cells, which in turn respond to infection or injury.

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.

One reason that HIV poses significant management challenges is because the virus directly targets T cells by gaining entry through a receptor. Once inside the cell, HIV then multiplies using the T cell's own genetic machinery. After the HIV virus replicates, it is transmitted directly from the infected T cell to macrophages. The presence of HIV can remain unrecognized for an extensive period of time before full disease symptoms develop.

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 21.8**b**. 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 21.8**a**. 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.

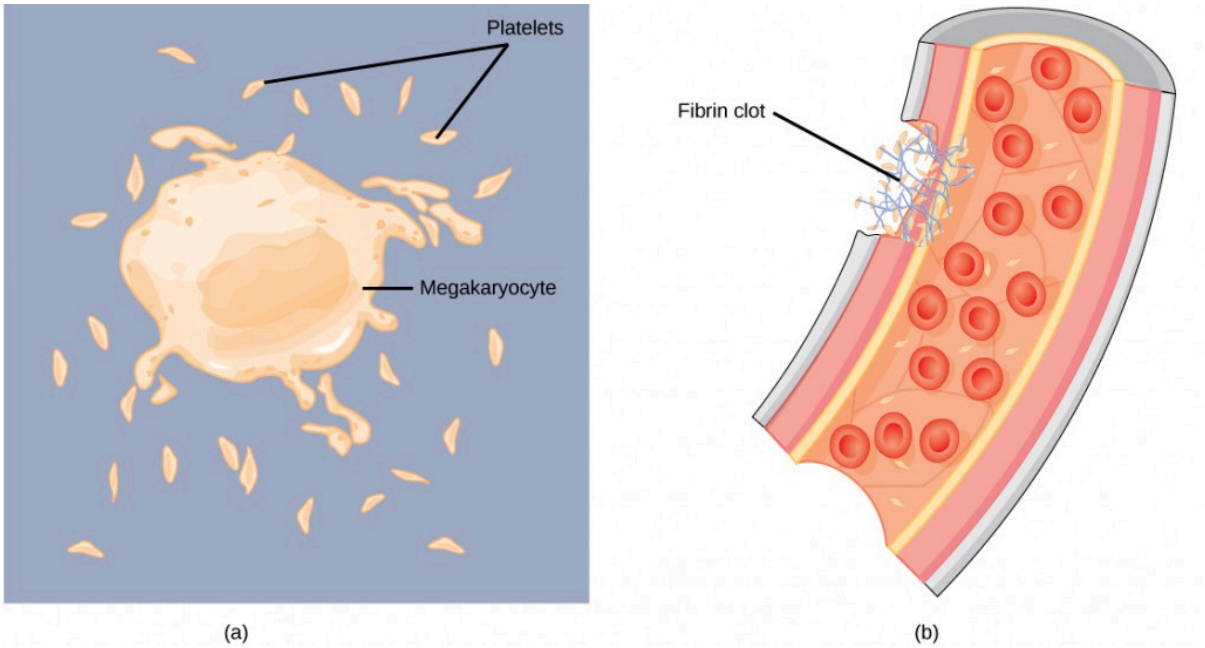


Figure 21.8. (a) Platelets are formed from large cells called megakaryocytes. The megakaryocyte breaks up into thousands of fragments that become platelets. (b) Platelets are required for clotting of the blood. The platelets collect at a wound site in conjunction with other clotting factors, such as 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 is separated 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.

Human serum albumin is the most abundant protein in human blood plasma and is synthesized in the liver. Albumin, which constitutes about half of the blood serum protein, transports hormones and fatty acids, buffers pH, and maintains osmotic pressures. Immunoglobin is a protein antibody produced in the mucosal lining and plays an important role in antibody mediated immunity.

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 21.9, 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 called agglutinougens are found in the blood plasma and react with the A or B antigens, if the two are mixed. 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. 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. The ABO blood groups were discovered in 1900 and 1901 by Karl Landsteiner at the University of Vienna.

The Rh blood group was first discovered in Rhesus monkeys. Most people have the Rh antigen (Rh+) and do not have anti-Rh antibodies in their blood. The few people who do not have the Rh antigen and are Rh– can develop anti-Rh antibodies if exposed to Rh+ blood. This can happen after a blood transfusion or after an Rh– woman has an Rh+ baby. The first exposure does not usually cause a reaction; however, at the second exposure, enough antibodies have built up in the blood to produce a reaction that causes agglutination and breakdown of red blood cells. An injection can prevent this reaction.

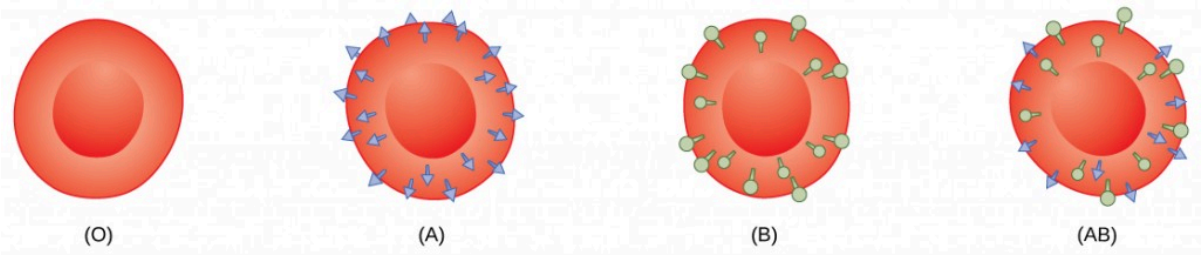


Figure 21.9. Human red blood cells may have either type A or B glycoproteins on their surface, both glycoproteins combined (AB), or neither (O). The glycoproteins serve as antigens and can elicit an immune response in a person who receives a transfusion containing unfamiliar antigens. Type O blood, which 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.

Concept in action

Play a blood typing game on the Nobel Prize website to solidify your understanding of blood types.

Summary

Specific components of the blood include red blood cells, white blood cells, platelets, and the plasma, which contains coagulation factors and serum. Blood is important for regulation of the body's pH, temperature, osmotic pressure, the circulation of nutrients and removal of waste, the distribution of hormones from endocrine glands, and the elimination of excess heat; it also contains components for blood clotting. Red blood cells are specialized cells that contain hemoglobin and circulate through the body delivering oxygen to cells. White blood cells are involved in the immune response to identify and target invading bacteria, viruses, and other foreign organisms; they also recycle waste components, such as old red blood cells. Platelets and blood clotting factors cause the change of the soluble protein fibrinogen to the insoluble protein fibrin at a wound site forming a plug. Plasma consists of 90 percent water along with various substances, such as coagulation factors and antibodies. The serum is the plasma component of the blood without the coagulation factors.

Exercises

1. White blood cells
    1. can be classified as granulocytes or agranulocytes
    2. defend the body against bacteria and viruses
    3. are also called leucocytes
    4. All of the above
2. Platelet plug formation occurs at which point?
    1. when large megakaryocytes break up into thousands of smaller fragments
    2. when platelets are dispersed through the blood stream
    3. when platelets are attracted to a site of blood vessel damage
    4. none of the above
3. In humans, the plasma comprises what percentage of the blood?
    1. 45 percent

2. 55 percent
3. 25 percent
4. 90 percent

4. The red blood cells of birds differ from mammalian red blood cells because:
    1. they are white and have nuclei
    2. they do not have nuclei
    3. they have nuclei
    4. they fight disease
5. Describe the cause of different blood type groups.
6. List some of the functions of blood in the body.
7. How does the lymphatic system work with blood flow?

Answers

1. D
2. C
3. B
4. C
5. Red blood cells are coated with proteins called antigens made of glycolipids and glycoproteins. When type A and type B blood are mixed, the blood agglutinates because of antibodies in the plasma that bind with the opposing antigen. Type O blood has no antigens. The Rh blood group has either the Rh antigen (Rh+) or no Rh antigen (Rh–).
6. Blood is important for regulation of the body's pH, temperature, and osmotic pressure, the circulation of nutrients and removal of wastes, the distribution of hormones from endocrine glands, the elimination of excess heat; it also contains components for the clotting of blood to prevent blood loss. Blood also transports clotting factors and disease-fighting agents.
7. Lymph capillaries take fluid from the blood to the lymph nodes. The lymph nodes filter the lymph by percolation through connective tissue filled with white blood cells. The white blood cells remove infectious agents, such as bacteria and viruses, to clean the lymph before it returns to the bloodstream.

Glossary

plasma
liquid component of blood that is left after the cells are removed

platelet
(also, thrombocyte) small cellular fragment that collects at wounds, cross-reacts with clotting factors, and forms a plug to prevent blood loss

red blood cell
small (7–8 μm) biconcave cell without mitochondria (and in mammals without nuclei) that is packed with hemoglobin, giving the cell its red color; transports oxygen through the body

serum
plasma without the coagulation factors

white blood cell
large (30 μm) cell with nuclei of which there are many types with different roles including the protection of the body from viruses and bacteria, and cleaning up dead cells and other waste

21.3. Mammalian Heart and Blood Vessels

Learning Objectives

By the end of this section, you will be able to:

- Describe the structure of the heart and explain how cardiac muscle is different from other muscles
- Describe the cardiac cycle
- Explain the structure of arteries, veins, and capillaries, and how blood flows through the body

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 21.10. Coronary circulation intrinsic to the heart takes blood directly from the main artery (aorta) coming from the heart. 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 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.

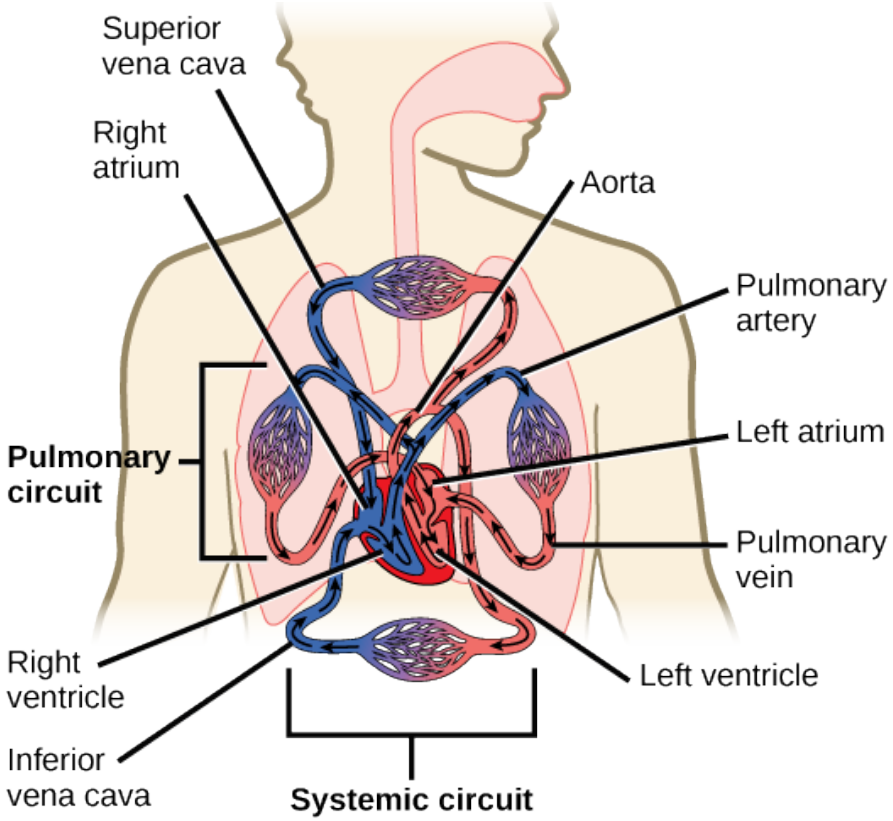


Figure 21.10. The mammalian circulatory system is divided into three circuits: the systemic circuit, the pulmonary circuit, and the coronary circuit. Blood is pumped from veins of the systemic circuit into the right atrium of the heart, then into the right ventricle. Blood then enters the pulmonary circuit, 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.

Which of the following statements about the circulatory system is false?

1. Blood in the pulmonary vein is deoxygenated.
2. Blood in the inferior vena cava is deoxygenated.
3. Blood in the pulmonary artery is deoxygenated.
4. Blood in the aorta is oxygenated.

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 21.11. 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, and the ventricles are the

chambers that pump blood. 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**, a flap of connective tissue that opens in only one direction to prevent the backflow of blood. The valve separating the chambers on the left side of the heart valve is called the biscuspid or mitral 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.

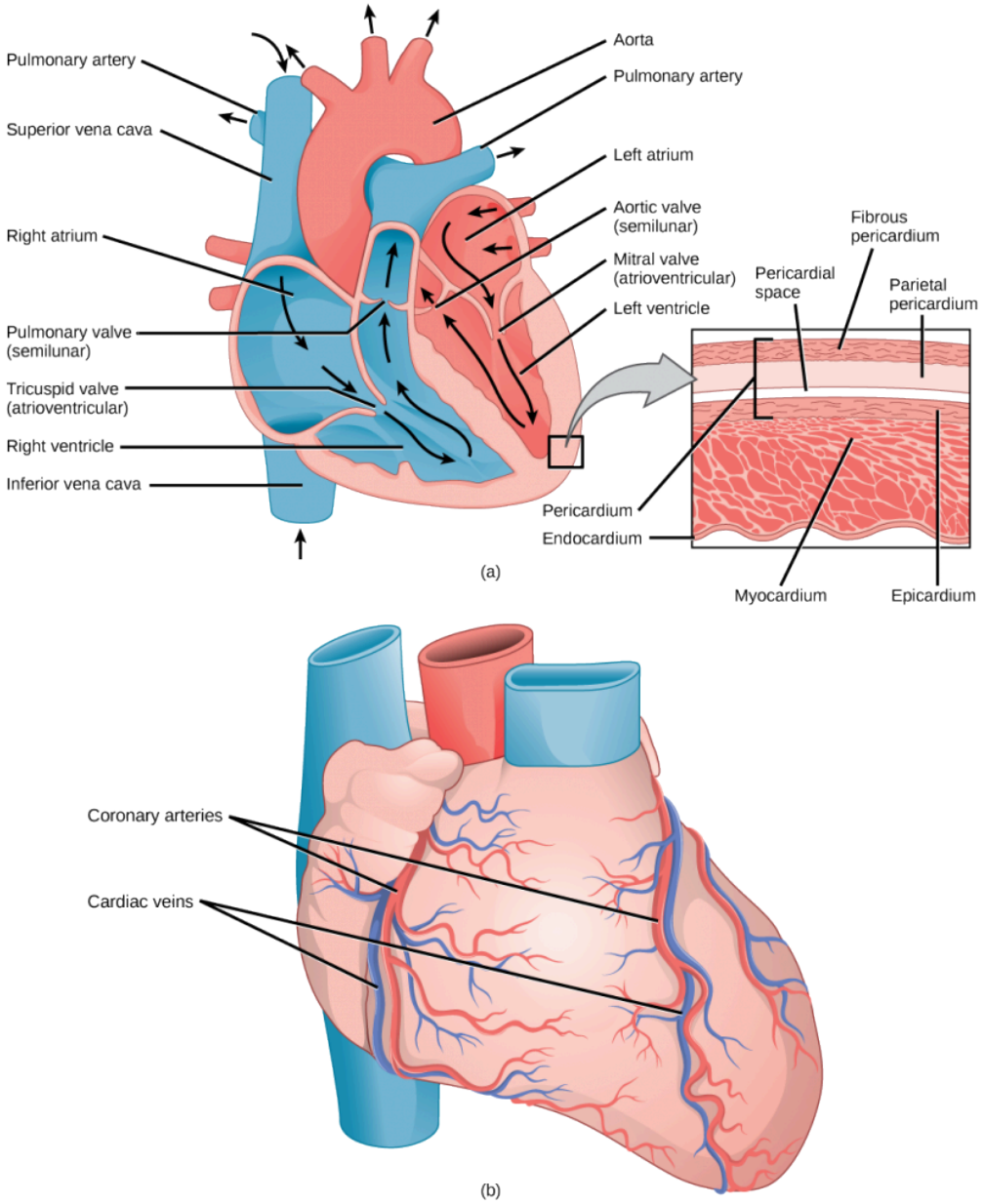


Figure 21.11. (a) The heart is primarily made of a thick muscle layer, called the myocardium, surrounded by membranes. One-way valves separate the four chambers. (b) Blood vessels of the coronary system, including the coronary arteries and veins, keep the heart musculature oxygenated.

Which of the following statements about the heart is false?

1. The mitral valve separates the left ventricle from the left atrium.
2. Blood travels through the bicuspid valve to the left atrium.
3. Both the aortic and the pulmonary valves are semilunar valves.
4. The mitral valve is an atrioventricular valve.

The heart is composed of three layers; the epicardium, the myocardium, and the endocardium,

illustrated in Figure 21.11. The inner wall of the heart has a lining called the **endocardium**. The **myocardium** consists of the heart muscle cells that make up the middle layer and the bulk of the heart wall. The outer layer of cells is called the **epicardium**, of which the second layer is a membranous layered structure called the **pericardium** that surrounds and protects the heart; it allows enough room for vigorous pumping but also keeps the heart in place to reduce friction between the heart and other structures.

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.

The Cardiac Cycle

The main purpose of the heart is to pump blood through the body; it does so in a repeating sequence called the cardiac cycle. The **cardiac cycle** is the coordination of the filling and emptying of the heart of blood by electrical signals that cause the heart muscles to contract and relax. The human heart beats over 100,000 times per day. In each cardiac cycle, the heart contracts (**systole**), pushing out the blood and pumping it through the body; this is followed by a relaxation phase (**diastole**), where the heart fills with blood, as illustrated in Figure 21.12. The atria contract at the same time, forcing blood through the atrioventricular valves into the ventricles. Closing of the atrioventricular valves produces a monosyllabic “lup” sound. Following a brief delay, the ventricles contract at the same time forcing blood through the semilunar valves into the aorta and the artery transporting blood to the lungs (via the pulmonary artery). Closing of the semilunar valves produces a monosyllabic “dup” sound.

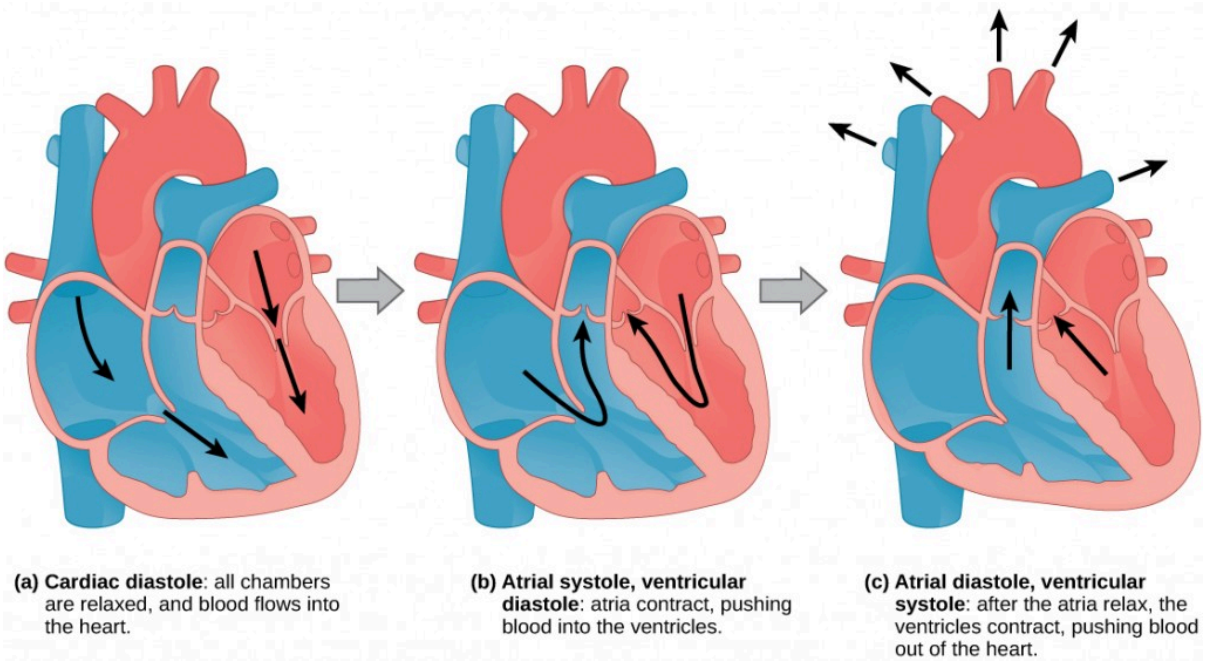


Figure 21.12. During (a) cardiac diastole, the heart muscle is relaxed and blood flows into the heart. During (b) atrial systole, the atria contract, pushing blood into the ventricles. During (c) atrial diastole, the ventricles contract, forcing blood out of the heart.

The pumping of the heart is a function of the cardiac muscle cells, or cardiomyocytes, that make up the heart muscle. **Cardiomyocytes**, shown in Figure 21.13, are distinctive muscle cells that are striated like skeletal muscle but pump rhythmically and involuntarily like smooth muscle; they are connected by intercalated disks exclusive to cardiac muscle. They are self-stimulated for a period of time and isolated cardiomyocytes will beat if given the correct balance of nutrients and electrolytes.

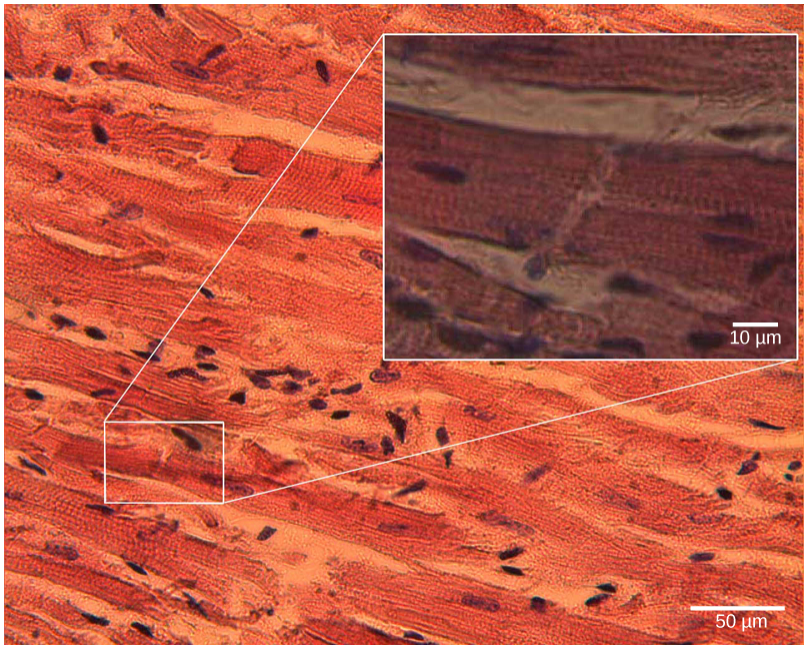


Figure 21.13. Cardiomyocytes are striated muscle cells found in cardiac tissue. (credit: modification of work by Dr. S. Girod, Anton Becker; scale-bar data from Matt Russell)

The autonomous beating of cardiac muscle cells is regulated by the heart's internal pacemaker that uses electrical signals to time the beating of the heart. The electrical signals and mechanical actions, illustrated in Figure 21.14, are intimately intertwined. The internal pacemaker starts at the **sinoatrial (SA) node**, which is located near the wall of the right atrium. Electrical charges spontaneously pulse from the SA node causing the two atria to contract in unison. The pulse reaches a second node, called the atrioventricular (AV) node, between the right atrium and right ventricle where it pauses for approximately 0.1 second before spreading to the walls of the ventricles. From the AV node, the electrical impulse enters the bundle of His, then to the left and right bundle branches extending through the interventricular septum. Finally, the Purkinje fibers conduct the impulse from the apex of the heart up the ventricular myocardium, and then the ventricles contract. This pause allows the atria to empty completely into the ventricles before the ventricles pump out the blood. The electrical impulses in the heart produce electrical currents that flow through the body and can be measured on the skin using electrodes. This information can be observed as an **electrocardiogram (ECG)**—a recording of the electrical impulses of the cardiac muscle.

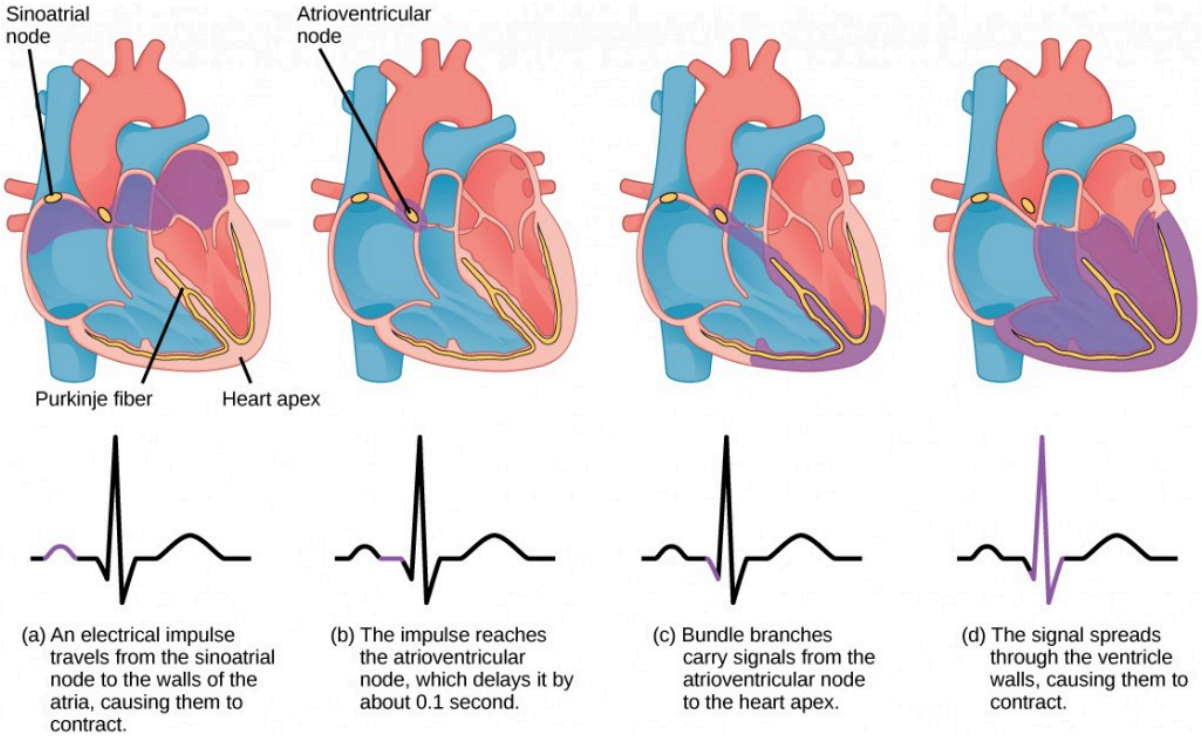


Figure 21.14. The beating of the heart is regulated by an electrical impulse that causes the characteristic reading of an ECG. The signal is initiated at the sinoatrial valve. The signal then (a) spreads to the atria, causing them to contract. The signal is (b) delayed at the atrioventricular node before it is passed on to the (c) heart apex. The delay allows the atria to relax before the (d) ventricles contract. The final part of the ECG cycle prepares the heart for the next beat.

Concept in Action

Visit this site to see the heart's "pacemaker" in action.

Arteries, Veins, and Capillaries

The blood from the heart is carried through the body by a complex network of blood vessels (Figure 21.15). **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.

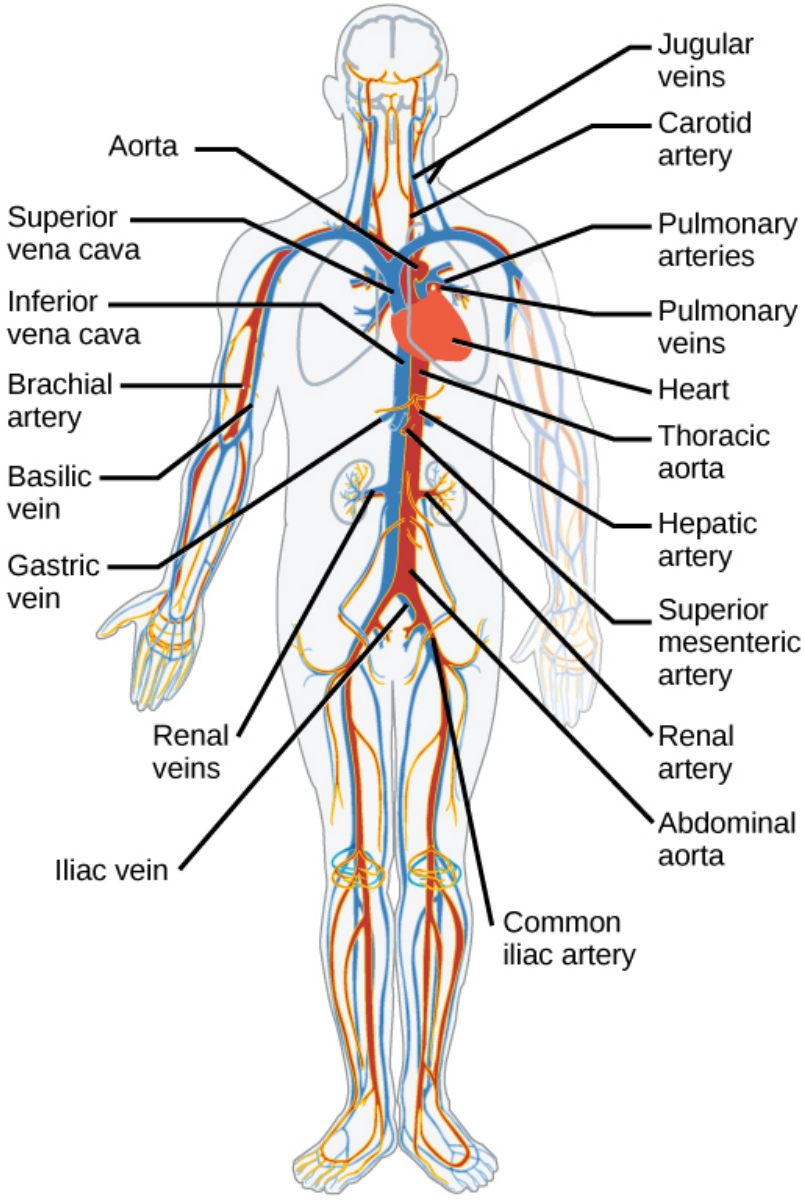


Figure 21.15. The major human arteries and veins are shown. (credit: modification of work by Mariana Ruiz Villareal)

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.

The structure of the different types of blood vessels reflects their function or layers. There are three distinct layers, or tunics, that form the walls of blood vessels (Figure 21.16). The first tunic is a smooth, inner lining of endothelial cells that are in contact with the red blood cells. The endothelial tunic is

continuous with the endocardium of the heart. In capillaries, this single layer of cells is the location of diffusion of oxygen and carbon dioxide between the endothelial cells and red blood cells, as well as the exchange site via endocytosis and exocytosis. The movement of materials at the site of capillaries is regulated by **vasoconstriction**, narrowing of the blood vessels, and **vasodilation,** widening of the blood vessels; this is important in the overall regulation of blood pressure.

Veins and arteries both have two further tunics that surround the endothelium: the middle tunic is composed of smooth muscle and the outermost layer is connective tissue (collagen and elastic fibers). The elastic connective tissue stretches and supports the blood vessels, and the smooth muscle layer helps regulate blood flow by altering vascular resistance through vasoconstriction and vasodilation. The arteries have thicker smooth muscle and connective tissue than the veins to accommodate the higher pressure and speed of freshly pumped blood. The veins are thinner walled as the pressure and rate of flow are much lower. In addition, veins are structurally different than arteries in that veins have valves to prevent the backflow of blood. Because veins have to work against gravity to get blood back to the heart, contraction of skeletal muscle assists with the flow of blood back to the heart.

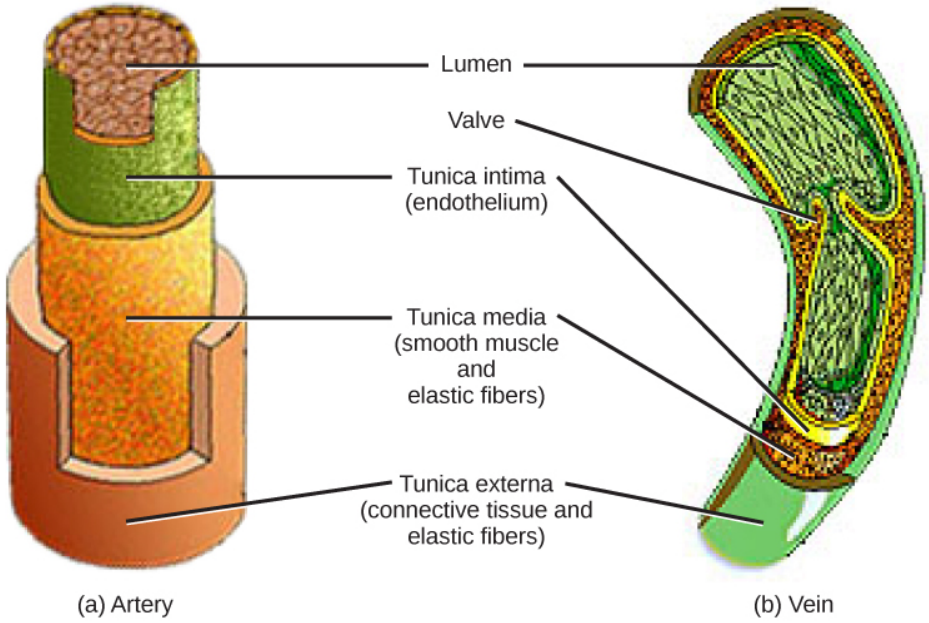


Figure 21.16. Arteries and veins consist of three layers: an outer tunica externa, a middle tunica media, and an inner tunica intima. Capillaries consist of a single layer of epithelial cells, the tunica intima. (credit: modification of work by NCI, NIH)

Summary

The heart muscle pumps blood through three divisions of the circulatory system: coronary, pulmonary, and systemic. There is one atrium and one ventricle on the right side and one atrium and one ventricle on the left side. The pumping of the heart is a function of cardiomyocytes, distinctive muscle cells that are striated like skeletal muscle but pump rhythmically and involuntarily like smooth muscle. The internal pacemaker starts at the sinoatrial node, which is located near the wall of the right atrium. Electrical charges pulse from the SA node causing the two atria to contract in unison; then the pulse reaches the atrioventricular node between the right atrium and right ventricle. A pause in the electric signal allows the atria to empty completely into the ventricles before the ventricles pump out the blood.

The blood from the heart is carried through the body by a complex network of blood vessels; arteries take blood away from the heart, and veins bring blood back to the heart.

Exercises

1. The heart's internal pacemaker beats by:
    1. an internal implant that sends an electrical impulse through the heart
    2. the excitation of cardiac muscle cells at the sinoatrial node followed by the atrioventricular node
    3. the excitation of cardiac muscle cells at the atrioventricular node followed by the sinoatrial node
    4. the action of the sinus
2. During the systolic phase of the cardiac cycle, the heart is _____.
    1. contracting
    2. relaxing
    3. contracting and relaxing
    4. filling with blood
3. Cardiomyocytes are similar to skeletal muscle because:
    1. they beat involuntarily
    2. they are used for weight lifting
    3. they pulse rhythmically
    4. they are striated
4. How do arteries differ from veins?
    1. Arteries have thicker smooth muscle layers to accommodate the changes in pressure from the heart.
    2. Arteries carry blood.
    3. Arteries have thinner smooth muscle layers and valves and move blood by the action of skeletal muscle.
    4. Arteries are thin walled and are used for gas exchange.
5. Describe the cardiac cycle.
6. What happens in capillaries?

Answers

1. B
2. A
3. D

4. A
5. The heart receives an electrical signal from the sinoatrial node triggering the cardiac muscle cells in the atria to contract. The signal pauses at the atrioventricular node before spreading to the walls of the ventricles so the blood is pumped through the body. This is the systolic phase. The heart then relaxes in the diastole and fills again with blood.
6. The capillaries basically exchange materials with their surroundings. Their walls are very thin and are made of one or two layers of cells, where gases, nutrients, and waste are diffused. They are distributed as beds, complex networks that link arteries as well as veins.

Glossary

angina
pain caused by partial blockage of the coronary arteries by the buildup of plaque and lack of oxygen to the heart muscle

aorta
major artery of the body that takes blood away from the heart

arteriole
small vessel that connects an artery to a capillary bed

artery
blood vessel that takes blood away from the heart

atherosclerosis
buildup of fatty plaques in the coronary arteries in the heart

bicuspid valve
(also, mitral valve; left atrioventricular valve) one-way membranous flap between the atrium and the ventricle in the left side of the heart

capillary bed
large number of capillaries that converge to take blood to a particular organ or tissue

capillary
smallest blood vessel that allows the passage of individual blood cells and the site of diffusion of oxygen and nutrient exchange

cardiac cycle
filling and emptying the heart of blood by electrical signals that cause the heart muscles to contract and relax

cardiac output
the volume of blood pumped by the heart in one minute as a product of heart rate multiplied by stroke volume

cardiomyocyte
specialized heart muscle cell that is striated but contracts involuntarily like smooth muscle

coronary artery
vessel that supplies the heart tissue with blood

coronary vein
vessel that takes blood away from the heart tissue back to the chambers in the heart

diastole
relaxation phase of the cardiac cycle when the heart is relaxed and the ventricles are filling with blood

electrocardiogram (ECG)
recording of the electrical impulses of the cardiac muscle

endocardium
innermost layer of tissue in the heart

epicardium
outermost tissue layer of the heart

inferior vena cava
drains blood from the veins that come from the lower organs and the legs

myocardial infarction
(also, heart attack) complete blockage of the coronary arteries and death of the cardiac muscle tissue

myocardium
heart muscle cells that make up the middle layer and the bulk of the heart wall

pericardium
membrane layer protecting the heart; also part of the epicardium

semilunar valve
membranous flap of connective tissue between the aorta and a ventricle of the heart (the aortic or pulmonary semilunar valves)

sinoatrial (SA) node
the heart's internal pacemaker; located near the wall of the right atrium

superior vena cava
drains blood from the jugular vein that comes from the brain and from the veins that come from the arms

systole
contraction phase of cardiac cycle when the ventricles are pumping blood into the arteries

tricuspid valve
one-way membranous flap of connective tissue between the atrium and the ventricle in the right side of the heart; also known as atrioventricular valve

vasoconstriction
narrowing of a blood vessel

vasodilation
widening of a blood vessel

21.4. Blood Flow and Blood Pressure Regulation

Learning Objectives

By the end of this section, you will be able to:

- Describe the system of blood flow through the body
- Describe how blood pressure is regulated

Blood pressure (BP) is the pressure exerted by blood on the walls of a blood vessel that helps to push blood through the body. Systolic blood pressure measures the amount of pressure that blood exerts on vessels while the heart is beating. The optimal systolic blood pressure is 120 mmHg. Diastolic blood pressure measures the pressure in the vessels between heartbeats. The optimal diastolic blood pressure is 80 mmHg. Many factors can affect blood pressure, such as hormones, stress, exercise, eating, sitting, and standing. Blood flow through the body is regulated by the size of blood vessels, by the action of smooth muscle, by one-way valves, and by the fluid pressure of the blood itself.

How Blood Flows Through the Body

Blood is pushed through the body by the action of the pumping heart. With each rhythmic pump, blood is pushed under high pressure and velocity away from the heart, initially along the main artery, the aorta. In the aorta, the blood travels at 30 cm/sec. As blood moves into the arteries, arterioles, and ultimately to the capillary beds, the rate of movement slows dramatically to about 0.026 cm/sec, one-thousand times slower than the rate of movement in the aorta. While the diameter of each individual arteriole and capillary is far narrower than the diameter of the aorta, and according to the law of continuity, fluid should travel faster through a narrower diameter tube, the rate is actually slower due to the overall diameter of all the combined capillaries being far greater than the diameter of the individual aorta.

The slow rate of travel through the capillary beds, which reach almost every cell in the body, assists with gas and nutrient exchange and also promotes the diffusion of fluid into the interstitial space. After the blood has passed through the capillary beds to the venules, veins, and finally to the main venae cavae, the rate of flow increases again but is still much slower than the initial rate in the aorta. Blood primarily moves in the veins by the rhythmic movement of smooth muscle in the vessel wall and by the action of the skeletal muscle as the body moves. Because most veins must move blood against the pull of gravity, blood is prevented from flowing backward in the veins by one-way valves. Because skeletal muscle contraction aids in venous blood flow, it is important to get up and move frequently after long periods of sitting so that blood will not pool in the extremities.

Blood flow through the capillary beds is regulated depending on the body’s needs and is directed by nerve and hormone signals. For example, after a large meal, most of the blood is diverted to the

stomach by vasodilation of vessels of the digestive system and vasoconstriction of other vessels. During exercise, blood is diverted to the skeletal muscles through vasodilation while blood to the digestive system would be lessened through vasoconstriction. The blood entering some capillary beds is controlled by small muscles, called precapillary sphincters, illustrated in Figure 21.17. If the sphincters are open, the blood will flow into the associated branches of the capillary blood. If all of the sphincters are closed, then the blood will flow directly from the arteriole to the venule through the thoroughfare channel (see Figure 21.17). These muscles allow the body to precisely control when capillary beds receive blood flow. At any given moment only about 5-10% of our capillary beds actually have blood flowing through them.

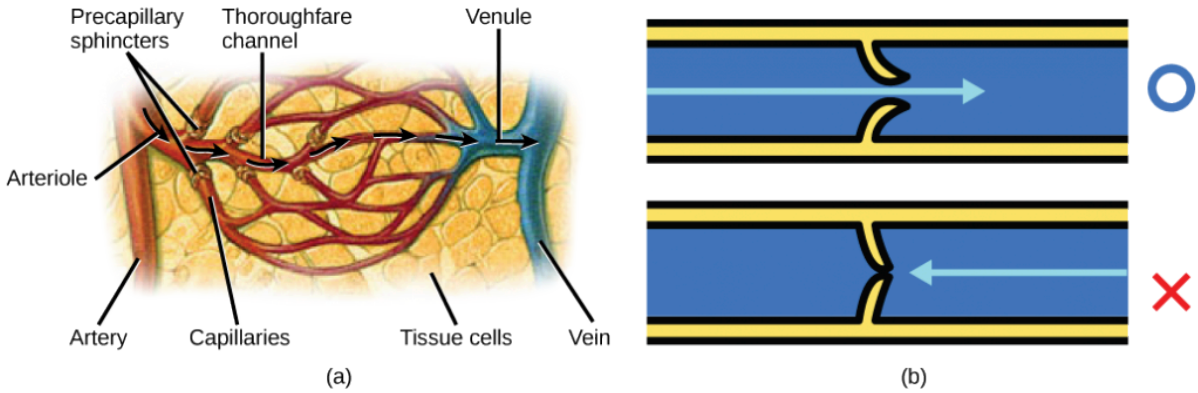


Figure 21.17. (a) Precapillary sphincters are rings of smooth muscle that regulate the flow of blood through capillaries; they help control the location of blood flow to where it is needed. (b) Valves in the veins prevent blood from moving backward. (credit a: modification of work by NCI)

Varicose veins are veins that become enlarged because the valves no longer close properly, allowing blood to flow backward. Varicose veins are often most prominent on the legs. Why do you think this is the case?

Concept in Action

Visit this site to see the circulatory system's blood flow.

Proteins and other large solutes cannot leave the capillaries. The loss of the watery plasma creates a hyperosmotic solution within the capillaries, especially near the venules. This causes about 85% of the plasma that leaves the capillaries to eventually diffuses back into the capillaries near the venules. The remaining 15% of blood plasma drains out from the interstitial fluid into nearby lymphatic vessels (Figure 21.18). The fluid in the lymph is similar in composition to the interstitial fluid. The lymph fluid passes through lymph nodes before it returns to the heart via the vena cava. **Lymph nodes** are specialized organs that filter the lymph by percolation through a maze of connective tissue filled with

white blood cells. The white blood cells remove infectious agents, such as bacteria and viruses, to clean the lymph before it returns to the bloodstream. After it is cleaned, the lymph returns to the heart by the action of smooth muscle pumping, skeletal muscle action, and one-way valves joining the returning blood near the junction of the venae cavae entering the right atrium of the heart.

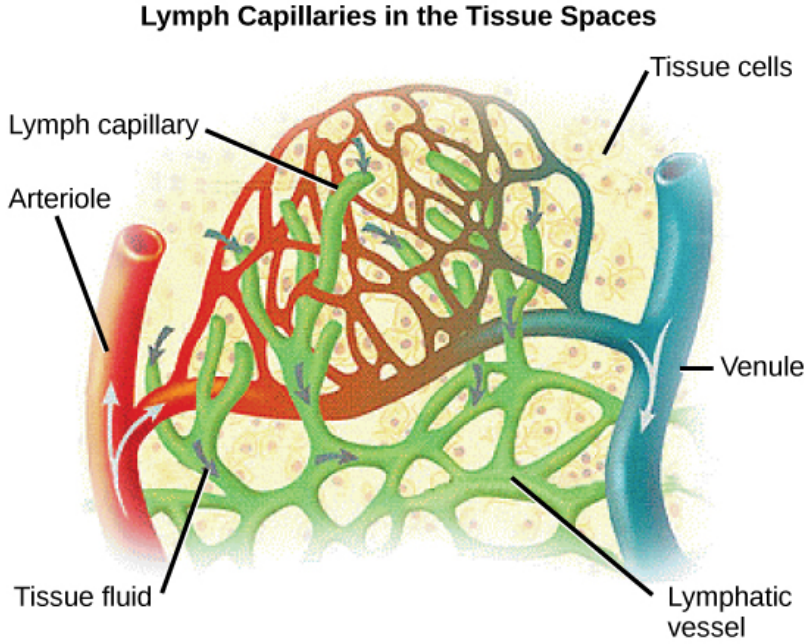


Figure 21.18. Fluid from the capillaries moves into the interstitial space and lymph capillaries by diffusion down a pressure gradient and also by osmosis. Out of 7,200 liters of fluid pumped by the average heart in a day, over 1,500 liters is filtered. (credit: modification of work by NCI, NIH)

Vertebrate Diversity in Blood Circulation

Blood circulation has evolved differently in vertebrates and may show variation in different animals for the required amount of pressure, organ and vessel location, and organ size. Animals with longs necks and those that live in cold environments have distinct blood pressure adaptations.

Long necked animals, such as giraffes, need to pump blood upward from the heart against gravity. The blood pressure required from the pumping of the left ventricle would be equivalent to 250 mm Hg (mm Hg = millimeters of mercury, a unit of pressure) to reach the height of a giraffe's head, which is 2.5 meters higher than the heart. However, if checks and balances were not in place, this blood pressure would damage the giraffe's brain, particularly if it was bending down to drink. These checks and balances include valves and feedback mechanisms that reduce the rate of cardiac output. Long-necked dinosaurs such as the sauropods had to pump blood even higher, up to ten meters above the heart. This would have required a blood pressure of more than 600 mm Hg, which could only have been achieved by an enormous heart. Evidence for such an enormous heart does not exist and mechanisms to reduce the blood pressure required include the slowing of metabolism as these animals grew larger. It is likely that they did not routinely feed on tree tops but grazed on the ground.

Living in cold water, whales need to maintain the temperature in their blood. This is achieved by the

veins and arteries being close together so that heat exchange can occur. This mechanism is called a countercurrent heat exchanger. The blood vessels and the whole body are also protected by thick layers of blubber to prevent heat loss. In land animals that live in cold environments, thick fur and hibernation are used to retain heat and slow metabolism.

Blood Pressure

The pressure of the blood flow in the body is produced by the hydrostatic pressure of the fluid (blood) against the walls of the blood vessels. Fluid will move from areas of high to low hydrostatic pressures. In the arteries, the hydrostatic pressure near the heart is very high and blood flows to the arterioles where the rate of flow is slowed by the narrow openings of the arterioles. During systole, when new blood is entering the arteries, the artery walls stretch to accommodate the increase of pressure of the extra blood; during diastole, the walls return to normal because of their elastic properties. The blood pressure of the systole phase and the diastole phase, graphed in Figure 21.19, gives the two pressure readings for blood pressure. For example, 120/80 indicates a reading of 120 mm Hg during the systole and 80 mm Hg during diastole. Throughout the cardiac cycle, the blood continues to empty into the arterioles at a relatively even rate. This resistance to blood flow is called **peripheral resistance.**

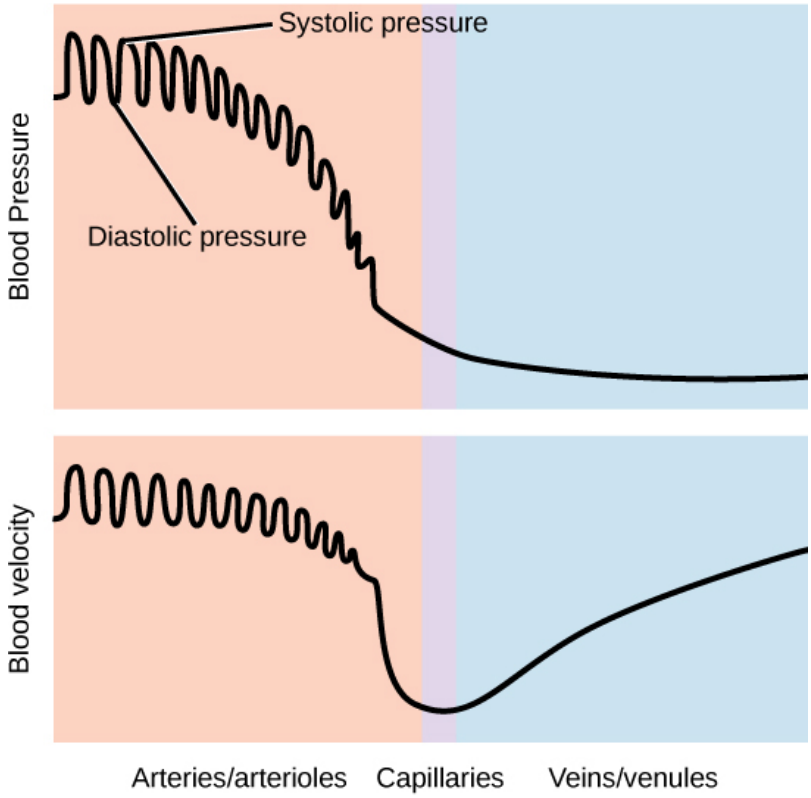


Figure 21.19. Blood pressure is related to the blood velocity in the arteries and arterioles. In the capillaries and veins, the blood pressure continues to decease but velocity increases.

Blood Pressure Regulation

Cardiac output is the volume of blood pumped by the heart in one minute. It is calculated by

multiplying the number of heart contractions that occur per minute (heart rate) times the **stroke volume** (the volume of blood pumped into the aorta per contraction of the left ventricle). Therefore, cardiac output can be increased by increasing heart rate, as when exercising. However, cardiac output can also be increased by increasing stroke volume, such as if the heart contracts with greater strength. Stroke volume can also be increased by speeding blood circulation through the body so that more blood enters the heart between contractions. During heavy exertion, the blood vessels relax and increase in diameter, offsetting the increased heart rate and ensuring adequate oxygenated blood gets to the muscles. Stress triggers a decrease in the diameter of the blood vessels, consequently increasing blood pressure. These changes can also be caused by nerve signals or hormones, and even standing up or lying down can have a great effect on blood pressure.

Summary

Blood primarily moves through the body by the rhythmic movement of smooth muscle in the vessel wall and by the action of the skeletal muscle as the body moves. Blood is prevented from flowing backward in the veins by one-way valves. Blood flow through the capillary beds is controlled by precapillary sphincters to increase and decrease flow depending on the body's needs and is directed by nerve and hormone signals. Lymph vessels take fluid that has leaked out of the blood to the lymph nodes where it is cleaned before returning to the heart. During systole, blood enters the arteries, and the artery walls stretch to accommodate the extra blood. During diastole, the artery walls return to normal. The blood pressure of the systole phase and the diastole phase gives the two pressure readings for blood pressure.

Exercises

1. Varicose veins are veins that become enlarged because the valves no longer close properly, allowing blood to flow backward. Varicose veins are often most prominent on the legs. Why do you think this is the case?
2. High blood pressure would be a result of _____.
    1. a high cardiac output and high peripheral resistance
    2. a high cardiac output and low peripheral resistance
    3. a low cardiac output and high peripheral resistance
    4. a low cardiac output and low peripheral resistance
3. How does blood pressure change during heavy exercise?

Answers

1. Blood in the legs is farthest away from the heart and has to flow up to reach it.
2. A
3. The heart rate increases, which increases the hydrostatic pressure against the artery walls. At the same time, the arterioles dilate in response to the increased exercise, which reduces peripheral resistance.

Glossary

blood pressure (BP)
pressure of blood in the arteries that helps to push blood through the body

lymph node
specialized organ that contains a large number of macrophages that clean the lymph before the fluid is returned to the heart

peripheral resistance
resistance of the artery and blood vessel walls to the pressure placed on them by the force of the heart pumping

stroke volume
– the volume of blood pumped into the aorta per contraction of the left ventricle

Chapter 22. Osmotic Regulation and Excretion

Figure 22.1. Just as humans recycle what we can and dump the remains into landfills, our bodies use and recycle what they can and excrete the remaining waste products. Our bodies' complex systems have developed ways to treat waste and maintain a balanced internal environment. (credit: modification of work by Redwin Law)

Introduction

The daily intake recommendation for human water consumption is eight to ten glasses of water. In order to achieve a healthy balance, the human body should excrete the eight to ten glasses of water every day. This occurs via the processes of urination, defecation, sweating and, to a small extent, respiration. The organs and tissues of the human body are soaked in fluids that are maintained at constant temperature, pH, and solute concentration, all crucial elements of homeostasis. The solutes in body fluids are mainly mineral salts and sugars, and osmotic regulation is the process by which the mineral salts and water are kept in balance. Osmotic homeostasis is maintained despite the influence of external factors like temperature, diet, and weather conditions.

22.1. Osmoregulation and Osmotic Balance

shrink due to water lossLearning Objectives

By the end of this section, you will be able to:

- Define osmosis and explain its role within molecules
- Explain why osmoregulation and osmotic balance are important body functions
- Describe active transport mechanisms
- Explain osmolarity and the way in which it is measured
- Describe osmoregulators or osmoconformers and how these tools allow animals to adapt to different environments

Osmosis is the diffusion of water across a membrane in response to **osmotic pressure** caused by an imbalance of molecules on either side of the membrane. **Osmoregulation** is the process of maintenance of salt and water balance (**osmotic balance**) across membranes within the body's fluids, which are composed of water, plus electrolytes and non-electrolytes. An **electrolyte** is a solute that dissociates into ions when dissolved in water. A **non-electrolyte**, in contrast, doesn't dissociate into ions during water dissolution. Both electrolytes and non-electrolytes contribute to the osmotic balance. The body's fluids include blood plasma, the cytosol within cells, and interstitial fluid, the fluid that exists in the spaces between cells and tissues of the body. The membranes of the body (such as the pleural, serous, and cell membranes) are **semi-permeable membranes**. Semi-permeable membranes are permeable (or permissive) to certain types of solutes and water. Solutions on two sides of a semi-permeable membrane tend to equalize in solute concentration by movement of solutes and/or water across the membrane. As seen in Figure 22.2, a cell placed in water tends to swell due to gain of water from the hypotonic or "low salt" environment. A cell placed in a solution with higher salt concentration, on the other hand, tends to make the membrane shrivel up due to loss of water into the hypertonic or "high salt" environment. Isotonic cells have an equal concentration of solutes inside and outside the cell; this equalizes the osmotic pressure on either side of the cell membrane which is a semi-permeable membrane.

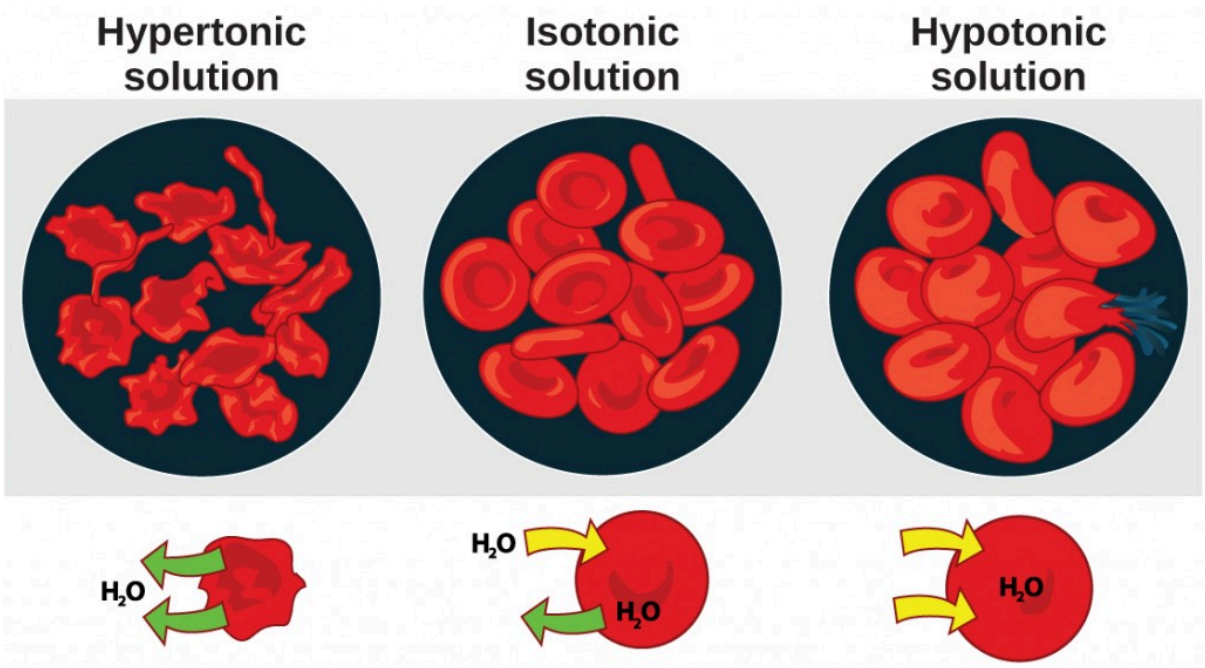


Figure 22.2. Cells placed in a hypertonic environment tend to shrink due to loss of water. In a hypotonic environment, cells tend to swell due to intake of water. The blood maintains an isotonic environment so that cells neither shrink nor swell. (credit: Mariana Ruiz Villareal)

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

Need for Osmoregulation

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

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

Transport of Electrolytes across Cell 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 number of solute atoms or molecules and not dependent on the size of the solute molecules. Because electrolytes dissociate into their component ions, they, in essence, add more solute particles into the solution and have a greater effect on osmotic pressure, per mass than compounds that do not dissociate in water, such as glucose.

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

Concept of Osmolality and Milliequivalent

In order to calculate osmotic pressure, it is necessary to understand how solute concentrations are measured. The unit for measuring solutes is the **mole**. One mole is defined as the gram molecular weight of the solute. For example, the molecular weight of sodium chloride is 58.44. Thus, one mole of sodium chloride weighs 58.44 grams. The **molarity** of a solution is the number of moles of solute per liter of solution. The **molality** of a solution is the number of moles of solute per kilogram of solvent. If the solvent is water, one kilogram of water is equal to one liter of water. While molarity and molality are used to express the concentration of solutions, electrolyte concentrations are usually expressed in terms of milliequivalents per liter (mEq/L): the mEq/L is equal to the ion concentration (in millimoles) multiplied by the number of electrical charges on the ion. The unit of milliequivalent takes into consideration the ions present in the solution (since electrolytes form ions in aqueous solutions) and the charge on the ions.

Thus, for ions that have a charge of one, one milliequivalent is equal to one millimole. For ions that have a charge of two (like calcium), one milliequivalent is equal to 0.5 millimoles. Another unit for the expression of electrolyte concentration is the milliosmole (mOsm), which is the number of milliequivalents of solute per kilogram of solvent. Body fluids are usually maintained within the range of 280 to 300 mOsm.

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 and molly that can tolerate a relatively wide range of salinity are referred to as euryhaline organisms. This is possible because some fish have evolved **osmoregulatory** mechanisms to survive in all kinds of aquatic environments. When they live in fresh water, their bodies tend to take up water because the environment is relatively hypotonic, as illustrated in Figure 22.3**a**. 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 22.3**b**. Most marine invertebrates, on the other hand, may be isotonic with sea water (**osmoconformers**). Their body fluid concentrations conform to changes in seawater concentration. Cartilaginous fishes' salt composition of the blood is similar to bony fishes; however, the blood of sharks contains the organic compounds urea and trimethylamine oxide (TMAO). This does not mean that their electrolyte composition is similar to that of 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 are cartilaginous fish with a rectal gland to secrete salt and assist in osmoregulation.

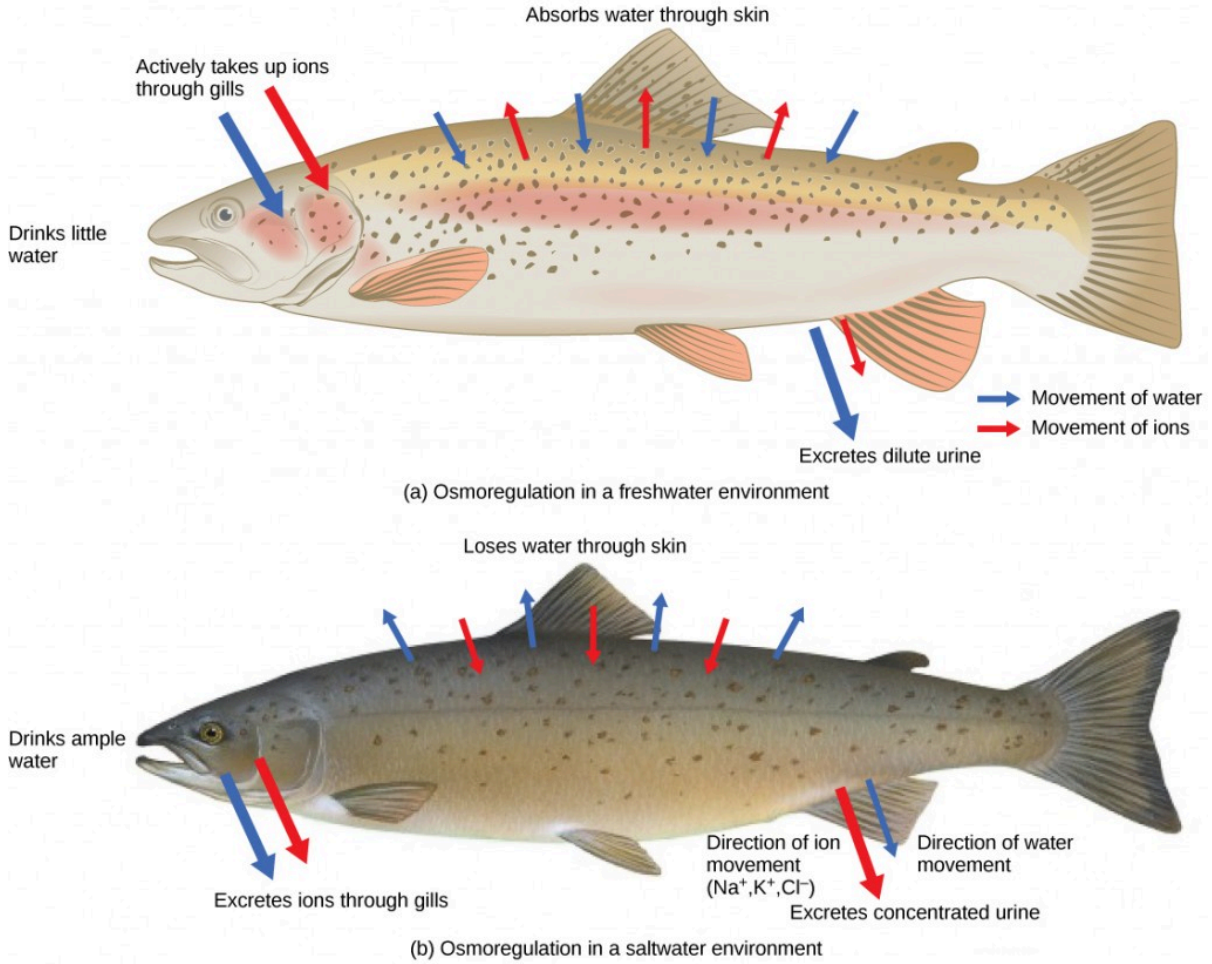


Figure 22.3. Fish are osmoregulators, but must use different mechanisms to survive in (a) freshwater or (b) saltwater environments. (credit: modification of work by Duane Raver, NOAA)

Dialysis Technician

Dialysis is a medical process of removing wastes and excess water from the blood by diffusion and ultrafiltration. When kidney function fails, dialysis must be done to artificially rid the body of wastes. This is a vital process to keep patients alive. In some cases, the patients undergo artificial dialysis until they are eligible for a kidney transplant. In others who are not candidates for kidney transplants, dialysis is a life-long necessity.

Dialysis technicians typically work in hospitals and clinics. While some roles in this field include equipment development and maintenance, most dialysis technicians work in direct patient care. Their on-the-job duties, which typically occur under the direct supervision of a registered nurse, focus on providing dialysis treatments. This can include reviewing patient history and current condition, assessing and responding to patient needs before and during treatment, and monitoring the dialysis process. Treatment may include taking and reporting a patient's vital signs and preparing solutions and equipment to ensure accurate and sterile procedures.

Summary

Solute concentrations across a semi-permeable membranes influence the movement of water and solutes across the membrane. It is the number of solute molecules and not the molecular size that is important in osmosis. Osmoregulation and osmotic balance are important bodily functions, resulting in water and salt balance. Not all solutes can pass through a semi-permeable membrane. Osmosis is the movement of water across the membrane. Osmosis occurs to equalize the number of solute molecules across a semi-permeable membrane by the movement of water to the side of higher solute concentration. Facilitated diffusion utilizes protein channels to move solute molecules from areas of higher to lower concentration while active transport mechanisms are required to move solutes against concentration gradients. Osmolarity is measured in units of milliequivalents or milliosmoles, both of which take into consideration the number of solute particles and the charge on them. Fish that live in fresh water or saltwater adapt by being osmoregulators or osmoconformers.

Exercises

1. When a dehydrated human patient needs to be given fluids intravenously, he or she is given:
    1. water, which is hypotonic with respect to body fluids
    2. saline at a concentration that is isotonic with respect to body fluids
    3. glucose because it is a non-electrolyte
    4. blood
2. The sodium ion is at the highest concentration in:
    1. intracellular fluid
    2. extracellular fluid
    3. blood plasma

4. none of the above

3. Cells in a hypertonic solution tend to:
    1. shrink due to water loss
    2. swell due to water gain
    3. stay the same size due to water moving into and out of the cell at the same rate
    4. none of the above
4. Why is excretion important in order to achieve osmotic balance?
5. Why do electrolyte ions move across membranes by active transport?

Answers

1. B
2. B
3. A
4. Excretion allows an organism to rid itself of waste molecules that could be toxic if allowed to accumulate. It also allows the organism to keep the amount of water and dissolved solutes in balance.
5. Electrolyte ions often require special mechanisms to cross the semi-permeable membranes in the body. Active transport is the movement against a concentration gradient.

Glossary

electrolyte
solute that breaks down into ions when dissolved in water

molality
number of moles of solute per kilogram of solvent

non-electrolyte
solute that does not break down into ions when dissolved in water

molarity
number of moles of solute per liter of solution

mole
gram equivalent of the molecular weight of a substance

osmoconformer
organism that changes its tonicity based on its environment

osmoregulation
mechanism by which water and solute concentrations are maintained at desired levels

osmoregulator
organism that maintains its tonicity irrespective of its environment

osmotic balance
balance of the amount of water and salt input and output to and from a biological system without disturbing the desired osmotic pressure and solute concentration in every compartment

osmotic pressure
pressure exerted on a membrane to equalize solute concentration on either side

semi-permeable membrane
membrane that allows only certain solutes to pass through

22.2. The Kidneys and Osmoregulatory Organs

Learning Objectives

By the end of this section, you will be able to:

- Explain how the kidneys serve as the main osmoregulatory organs in mammalian systems
- Describe the structure of the kidneys and the functions of the parts of the kidney
- Describe how the nephron is the functional unit of the kidney and explain how it actively filters blood and generates urine
- Detail the three steps in the formation of urine: glomerular filtration, tubular reabsorption, and tubular secretion

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

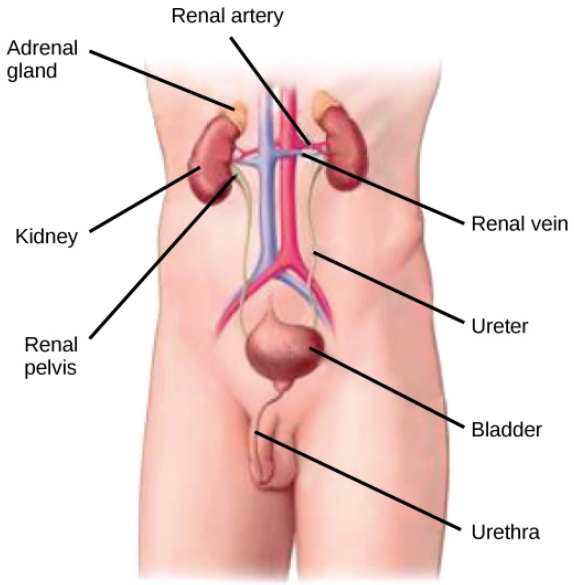


Figure 22.4. Kidneys filter the blood, producing urine that is stored in the bladder prior to elimination through the urethra. (credit: modification of work by NCI)

Kidney Structure

Externally, the kidneys are surrounded by three layers, illustrated in Figure 22.5. 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**.

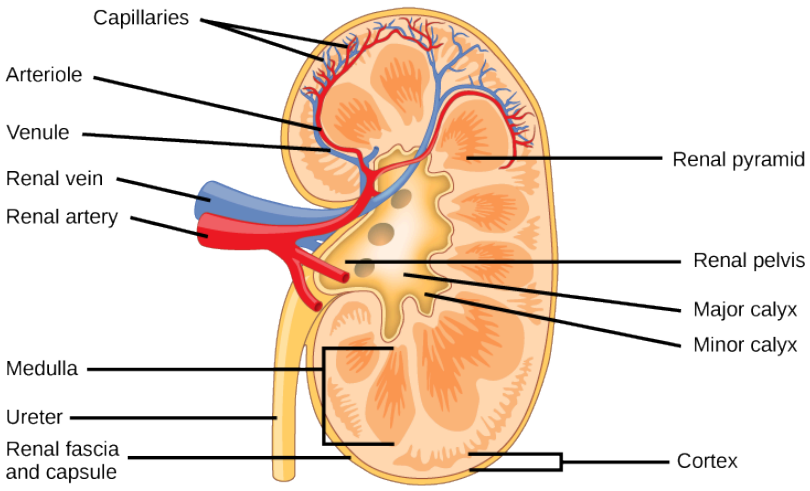


Figure 22.5. The internal structure of the kidney is shown. (credit: modification of work by NCI)

Which of the following statements about the kidney is false?

1. The renal pelvis drains into the ureter.
2. The renal pyramids are in the medulla.
3. The cortex covers the capsule.
4. Nephrons are in the renal cortex.

Because the kidney filters blood, its network of blood vessels is an important component of its structure and function. The arteries, veins, and nerves that supply the kidney enter and exit at the renal hilum. Renal blood supply starts with the branching of the aorta into the **renal arteries** (which are each named based on the region of the kidney they pass through) and ends with the exiting of the **renal veins** to join the **inferior vena cava**. The renal arteries split into several **segmental arteries** upon entering the kidneys. Each segmental artery splits further into several **interlobar arteries** and enters the renal columns, which supply the renal lobes. The interlobar arteries split at the junction of the renal cortex and medulla to form the **arcuate arteries**. The arcuate "bow shaped" arteries form arcs along the base of the medullary pyramids. **Cortical radiate arteries**, as the name suggests, radiate out from the arcuate arteries. The cortical radiate arteries branch into numerous afferent arterioles, and then enter the capillaries supplying the nephrons. Veins trace the path of the arteries and have similar names, except there are no segmental veins.

As mentioned previously, the functional unit of the kidney is the nephron, illustrated in Figure 22.6. Each kidney is made up of over one million nephrons that dot the renal cortex, giving it a granular appearance when sectioned sagittally. There are two types of nephrons— **cortical nephrons** (85 percent), which are deep in the renal cortex, and **juxtamedullary nephrons** (15 percent), which lie in the renal cortex close to the renal medulla. A nephron consists of three parts—a **renal corpuscle**, a **renal tubule**, and the associated capillary network, which originates from the cortical radiate arteries.

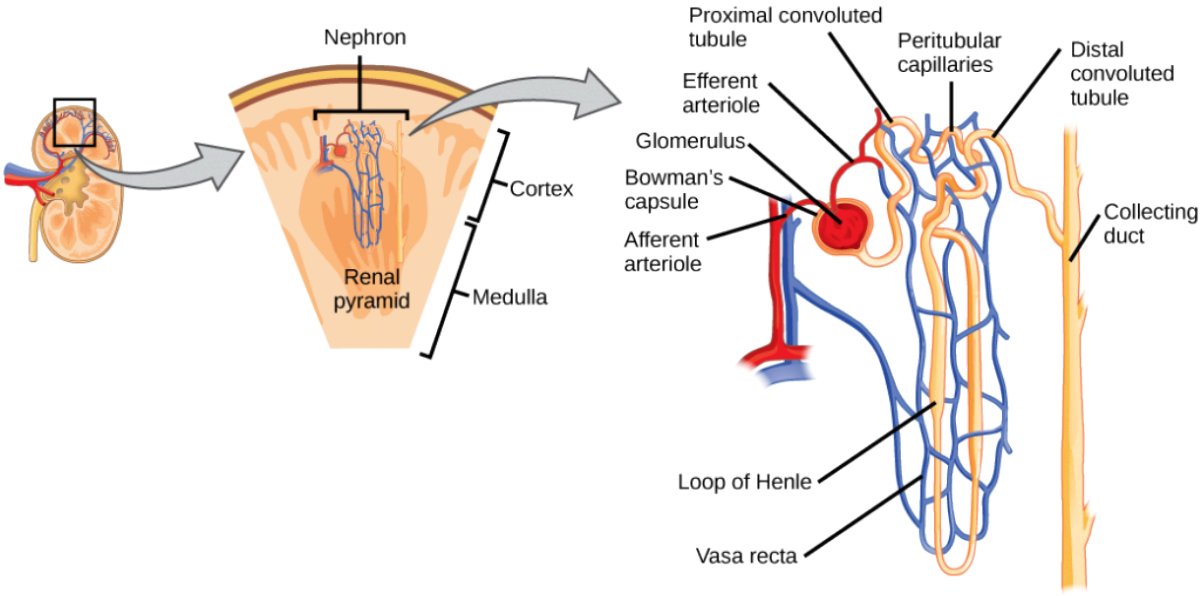


Figure 22.6. The nephron is the functional unit of the kidney. The glomerulus and convoluted tubules are located in the kidney cortex, while collecting ducts are located in the pyramids of the medulla. (credit: modification of work by NIDDK)

Which of the following statements about the nephron is false?

1. The collecting duct empties into the distal convoluted tubule.
2. The Bowman's capsule surrounds the glomerulus.
3. The loop of Henle is between the proximal and distal convoluted tubules.
4. The loop of Henle empties into the distal convoluted tubule.

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. In cortical nephrons, the peritubular capillary network surrounds the PCT and DCT. In juxtamedullary nephrons, the peritubular capillary network forms a network around the loop of Henle and is called the **vasa recta**.

Concept in Action

Go to this website to see another coronal section of the kidney and to explore an animation of the workings of nephrons.

Kidney Function and Physiology

Kidneys filter blood in a three-step process. First, the nephrons filter blood that runs through the capillary network in the glomerulus. Almost all solutes, except for proteins, are filtered out into the glomerulus by a process called **glomerular filtration**. Second, the filtrate is collected in the renal tubules. Most of the solutes get reabsorbed in the PCT by a process called **tubular reabsorption**. In the loop of Henle, the filtrate continues to exchange solutes and water with the renal medulla and the peritubular capillary network. Water is also reabsorbed during this step. Then, additional solutes and wastes are secreted into the kidney tubules during **tubular secretion**, which is, in essence, the opposite process 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 22.7.

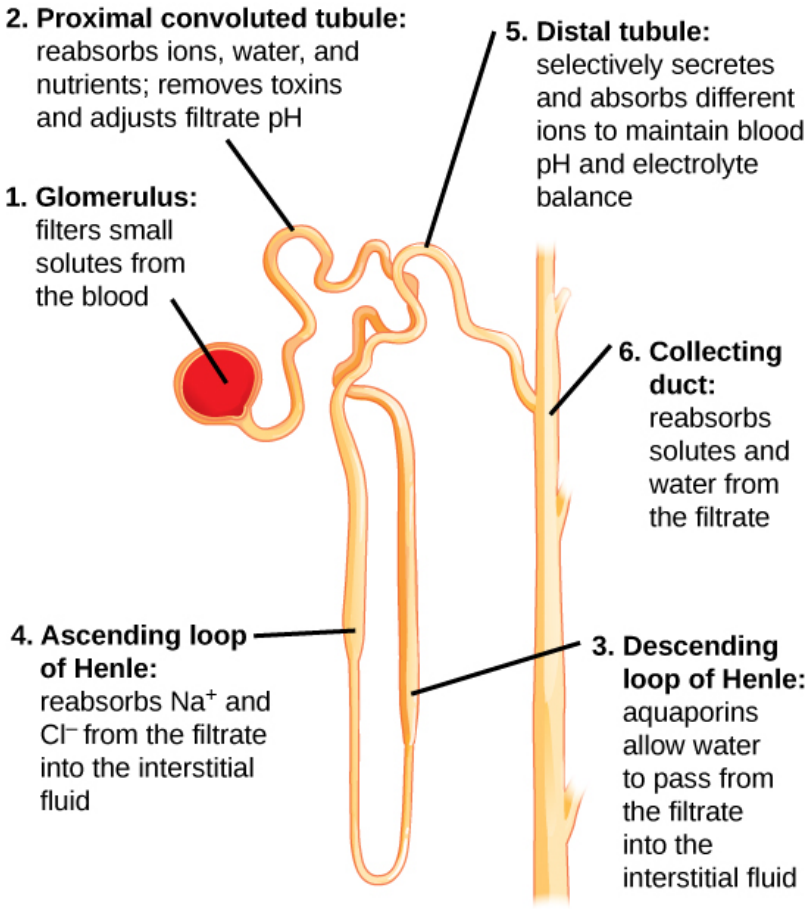


Figure 22.7. Each part of the nephron performs a different function in filtering waste and maintaining homeostatic balance. (1) The glomerulus forces small solutes out of the blood by pressure. (2) The proximal convoluted tubule reabsorbs ions, water, and nutrients from the filtrate into the interstitial fluid, and actively transports toxins and drugs from the interstitial fluid into the filtrate. The proximal convoluted tubule also adjusts blood pH by selectively secreting ammonia (NH3) into the filtrate, where it reacts with H+ to form NH4+. The more acidic the filtrate, the more ammonia is secreted. (3) The descending loop of Henle is lined with cells containing aquaporins that allow water to pass from the filtrate into the interstitial fluid. (4) In the thin part of the ascending loop of Henle, Na+ and Cl- ions diffuse into the interstitial fluid. In the thick part, these same ions are actively transported into the interstitial fluid. Because salt but not water is lost, the filtrate becomes more dilute as it travels up the limb. (5) In the distal convoluted tubule, K+ and H+ ions are selectively secreted into the filtrate, while Na+, Cl-, and HCO3- ions are reabsorbed to maintain pH and electrolyte balance in the blood. (6) The collecting duct reabsorbs solutes and water from the filtrate, forming dilute urine. (credit: modification of work by NIDDK)

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. **Glomerular filtration rate (GFR)** is the volume of glomerular filtrate formed per minute by the kidneys. GFR is regulated by multiple mechanisms and is an important indicator of kidney function.

Concept in Action

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

Tubular Reabsorption and Secretion

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

In the loop of Henle, the permeability of the membrane changes. The descending limb is permeable to water, not solutes; the opposite is true for the ascending limb. Additionally, the loop of Henle invades the renal medulla, which is naturally high in salt concentration and tends to absorb water from the renal tubule and concentrate the filtrate. The osmotic gradient increases as it moves deeper into the medulla. Because two sides of the loop of Henle perform opposing functions, as illustrated in Figure 22.8, it acts as a **countercurrent multiplier**. The vasa recta around it acts as the **countercurrent exchanger**.

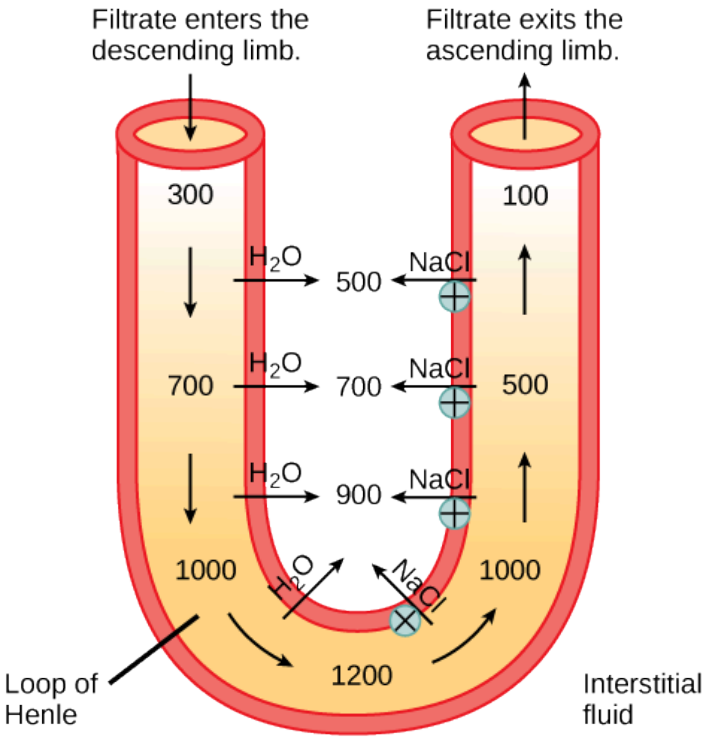


Figure 22.8. The loop of Henle acts as a countercurrent multiplier that uses energy to create concentration gradients. The descending limb is water permeable. Water flows from the filtrate to the interstitial fluid, so osmolality inside the limb increases as it descends into the renal medulla. At the bottom, the osmolality is higher inside the loop than in the interstitial fluid. Thus, as filtrate enters the ascending limb, Na+ and Cl- ions exit through ion channels present in the cell membrane. Further up, Na+ is actively transported out of the filtrate and Cl- follows. Osmolarity is given in units of milliosmoles per liter (mOsm/L).

Loop diuretics are drugs sometimes used to treat hypertension. These drugs 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.

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

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

Nephrologist

A nephrologist studies and deals with diseases of the kidneys—both those that cause kidney failure (such as diabetes) and the conditions that are produced by kidney disease (such as hypertension). Blood pressure, blood volume, and changes in electrolyte balance come under the purview of a nephrologist.

Nephrologists usually work with other physicians who refer patients to them or consult with them about specific diagnoses and treatment plans. Patients are usually referred to a nephrologist for symptoms such as blood or protein in the urine, very high blood pressure, kidney stones, or renal failure.

Nephrology is a subspecialty of internal medicine. To become a nephrologist, medical school is followed by additional training to become certified in internal medicine. An additional two or more years is spent specifically studying kidney disorders and their accompanying effects on the body.

Summary

The kidneys are the main osmoregulatory organs in mammalian systems; they function to filter blood and maintain the osmolarity of body fluids at 300 mOsm. They are surrounded by three layers and are made up internally of three distinct regions—the cortex, medulla, and pelvis.

The blood vessels that transport blood into and out of the kidneys arise from and merge with the aorta and inferior vena cava, respectively. The renal arteries branch out from the aorta and enter the kidney where they further divide into segmental, interlobar, arcuate, and cortical radiate arteries.

The nephron is the functional unit of the kidney, which actively filters blood and generates urine. The nephron is made up of the renal corpuscle and renal tubule. Cortical nephrons are found in the renal cortex, while juxtamedullary nephrons are found in the renal cortex close to the renal medulla. The nephron filters and exchanges water and solutes with two sets of blood vessels and the tissue fluid in the kidneys.

There are three steps in the formation of urine: glomerular filtration, which occurs in the glomerulus; tubular reabsorption, which occurs in the renal tubules; and tubular secretion, which also occurs in the renal tubules.

Exercises

1. Which of the following statements about the kidney is false?

1. The renal pelvis drains into the ureter.
2. The renal pyramids are in the medulla.
3. The cortex covers the capsule.
4. Nephrons are in the renal cortex.

2. Which of the following statements about the nephron is false?

    1. The collecting duct empties into the distal convoluted tubule.
    2. The Bowman's capsule surrounds the glomerulus.
    3. The loop of Henle is between the proximal and distal convoluted tubules.
    4. The loop of Henle empties into the distal convoluted tubule.

3. The macula densa is/are:

    1. present in the renal medulla.
    2. dense tissue present in the outer layer of the kidney.
    3. cells present in the DCT and collecting tubules.
    4. present in blood capillaries.

4. The osmolarity of body fluids is maintained at _____.

    1. 100 mOsm
    2. 300 mOsm
    3. 1000 mOsm
    4. it is not constantly maintained

5. The gland located at the top of the kidney is the _____ gland.

    1. adrenal
    2. pituitary
    3. thyroid
    4. thymus

6. Loop diuretics are drugs sometimes used to treat hypertension. These drugs 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?
7. Why are the loop of Henle and vasa recta important for the formation of concentrated urine?
8. Describe the structure of the kidney.

Answers

1. C
2. A
3. C

4. B
5. A
6. Loop diuretics decrease the excretion of salt into the renal medulla, thereby reducing its osmolality. As a result, less water is excreted into the medulla by the descending limb, and more water is excreted as urine.
7. The loop of Henle is part of the renal tubule that loops into the renal medulla. In the loop of Henle, the filtrate exchanges solutes and water with the renal medulla and the vasa recta (the peritubular capillary network). The vasa recta acts as the countercurrent exchanger. The kidneys maintain the osmolality of the rest of the body at a constant 300 mOsm by concentrating the filtrate as it passes through the loop of Henle.
8. Externally, the kidneys are surrounded by three layers. 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, which is the concave part of the “bean” shape.

Glossary

afferent arteriole
arteriole that branches from the cortical radiate artery and enters the glomerulus

arcuate artery
artery that branches from the interlobar artery and arches over the base of the renal pyramids

ascending limb
part of the loop of Henle that ascends from the renal medulla to the renal cortex

Bowman’s capsule
structure that encloses the glomerulus

calyx
structure that connects the renal pelvis to the renal medulla

cortex (animal)
outer layer of an organ like the kidney or adrenal gland

cortical radiate artery
artery that radiates from the arcuate arteries into the renal cortex

cortical nephron
nephron that lies in the renal cortex

countercurrent exchanger
peritubular capillary network that allows exchange of solutes and water from the renal tubules

countercurrent multiplier
osmotic gradient in the renal medulla that is responsible for concentration of urine

descending limb
part of the loop of Henle that descends from the renal cortex into the renal medulla

distal convoluted tubule (DCT)
part of the renal tubule that is the most distant from the glomerulus

efferent arteriole
arteriole that exits from the glomerulus

glomerular filtration
filtration of blood in the glomerular capillary network into the glomerulus

glomerular filtration rate (GFR)
amount of filtrate formed by the glomerulus per minute

glomerulus (renal)
part of the renal corpuscle that contains the capillary network

hilum
region in the renal pelvis where blood vessels, nerves, and ureters bunch before entering or exiting the kidney

inferior vena cava
one of the main veins in the human body

interlobar artery
artery that branches from the segmental artery and travels in between the renal lobes

juxtaglomerular cell
cell in the afferent and efferent arterioles that responds to stimuli from the macula densa

juxtamedullary nephron
nephron that lies in the cortex but close to the renal medulla

kidney
organ that performs excretory and osmoregulatory functions

lobes of the kidney
renal pyramid along with the adjoining cortical region

loop of Henle
part of the renal tubule that loops into the renal medulla

macula densa
group of cells that senses changes in sodium ion concentration; present in parts of the renal tubule and collecting ducts

medulla
middle layer of an organ like the kidney or adrenal gland

nephron
functional unit of the kidney

perirenal fat capsule
fat layer that suspends the kidneys

peritubular capillary network
capillary network that surrounds the renal tubule after the efferent artery exits the glomerulus

proximal convoluted tubule (PCT)
part of the renal tubule that lies close to the glomerulus

renal artery
branch of the artery that enters the kidney

renal capsule
layer that encapsulates the kidneys

renal column
area of the kidney through which the interlobar arteries travel in the process of supplying blood to the renal lobes

renal corpuscle
glomerulus and the Bowman's capsule together

renal fascia
connective tissue that supports the kidneys

renal pelvis
region in the kidney where the calyces join the ureters

renal pyramid
conical structure in the renal medulla

renal tubule
tubule of the nephron that arises from the glomerulus

renal vein
branch of a vein that exits the kidney and joins the inferior vena cava

segmental artery
artery that branches from the renal artery

transport maximum
maximum amount of solute that can be transported out of the renal tubules during reabsorption

tubular reabsorption
reclamation of water and solutes that got filtered out in the glomerulus

tubular secretion
process of secretion of wastes that do not get reabsorbed

ureter
urine-bearing tube coming out of the kidney; carries urine to the bladder

urinary bladder
structure that the ureters empty the urine into; stores urine

urine
filtrate produced by kidneys that gets excreted out of the body

vasa recta
peritubular network that surrounds the loop of Henle of the juxtamedullary nephrons

22.3. Excretion Systems

Learning Objectives

By the end of this section, you will be able to:

- Explain how vacuoles, present in microorganisms, work to excrete waste
- Describe the way in which flame cells and nephridia in worms perform excretory functions and maintain osmotic balance
- Explain how insects use Malpighian tubules to excrete wastes and maintain osmotic balance

Microorganisms and invertebrate animals use more primitive and simple mechanisms to get rid of their metabolic wastes than the mammalian system of kidney and urinary function. Three excretory systems evolved in organisms before complex kidneys: vacuoles, flame cells, and Malpighian tubules.

Contractile Vacuoles in Microorganisms

The most fundamental feature of life is the presence of a cell. In other words, a cell is the simplest functional unit of a life. Bacteria are unicellular, prokaryotic organisms that have some of the least complex life processes in place; however, prokaryotes such as bacteria do not contain membrane-bound vacuoles. The cells of microorganisms like bacteria, protozoa, and fungi are bound by cell membranes and use them to interact with the environment. Some cells, including some leucocytes in humans, are able to engulf food by endocytosis—the formation of vesicles by involution of the cell membrane within the cells. The same vesicles are able to interact and exchange metabolites with the intracellular environment. In some unicellular eukaryotic organisms such as the amoeba, shown in Figure 22.9, cellular wastes and excess water are excreted by exocytosis, when the contractile vacuoles merge with the cell membrane and expel wastes into the environment. Contractile vacuoles (CV) should not be confused with vacuoles, which store food or water.

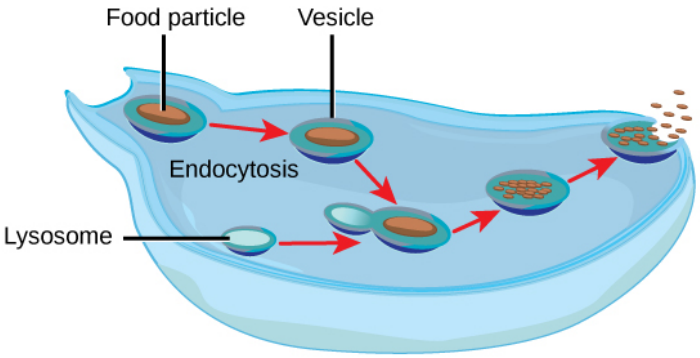


Figure 22.9. Some unicellular organisms, such as the amoeba, ingest food by endocytosis. The food vesicle fuses with a lysosome, which digests the food. Waste is excreted by exocytosis.

Flame Cells of Planaria and Nephridia of Worms

As multi-cellular systems evolved to have organ systems that divided the metabolic needs of the body, individual organs evolved to perform the excretory function. Planaria are flatworms that live in fresh water. Their excretory system consists of two tubules connected to a highly branched duct system. The cells in the tubules are called **flame cells** (or **protonephridia**) because they have a cluster of cilia that looks like a flickering flame when viewed under the microscope, as illustrated in Figure 22.10**a**. The cilia propel waste matter down the tubules and out of the body through excretory pores that open on the body surface; cilia also draw water from the interstitial fluid, allowing for filtration. Any valuable metabolites are recovered by reabsorption. Flame cells are found in flatworms, including parasitic tapeworms and free-living planaria. They also maintain the organism's osmotic balance.

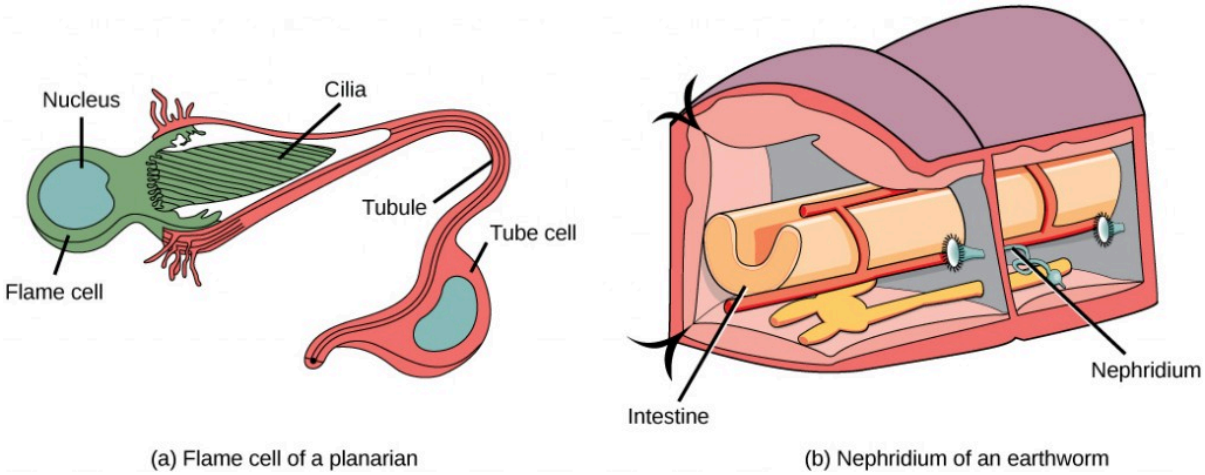


Figure 22.10. In the excretory system of the (a) planaria, cilia of flame cells propel waste through a tubule formed by a tube cell. Tubules are connected into branched structures that lead to pores located all along the sides of the body. The filtrate is secreted through these pores. In (b) annelids such as earthworms, nephridia filter fluid from the coelom, or body cavity. Beating cilia at the opening of the nephridium draw water from the coelom into a tubule. As the filtrate passes down the tubules, nutrients and other solutes are reabsorbed by capillaries. Filtered fluid containing nitrogenous and other wastes is stored in a bladder and then secreted through a pore in the side of the body.

Earthworms (annelids) have slightly more evolved excretory structures called **nephridia**, illustrated in

Figure 22.10**b**. A pair of nephridia is present on each segment of the earthworm. They are similar to flame cells in that they have a tubule with cilia. Excretion occurs through a pore called the **nephridiopore**. They are more evolved than the flame cells in that they have a system for tubular reabsorption by a capillary network before excretion.

Malpighian Tubules of Insects

Malpighian tubules are found lining the gut of some species of arthropods, such as the bee illustrated in Figure 22.11. They are usually found in pairs and the number of tubules varies with the species of insect. Malpighian tubules are convoluted, which increases their surface area, and they are lined with **microvilli** for reabsorption and maintenance of osmotic balance. Malpighian tubules work cooperatively with specialized glands in the wall of the rectum. Body fluids are not filtered as in the case of nephridia; urine is produced by tubular secretion mechanisms by the cells lining the Malpighian tubules that are bathed in hemolymph (a mixture of blood and interstitial fluid that is found in insects and other arthropods as well as most mollusks). Metabolic wastes like uric acid freely diffuse into the tubules. There are exchange pumps lining the tubules, which actively transport H^+ ions into the cell and K^+ or Na^+ ions out; water passively follows to form urine. The secretion of ions alters the osmotic pressure which draws water, electrolytes, and nitrogenous waste (uric acid) into the tubules. Water and electrolytes are reabsorbed when these organisms are faced with low-water environments, and uric acid is excreted as a thick paste or powder. Not dissolving wastes in water helps these organisms to conserve water; this is especially important for life in dry environments.

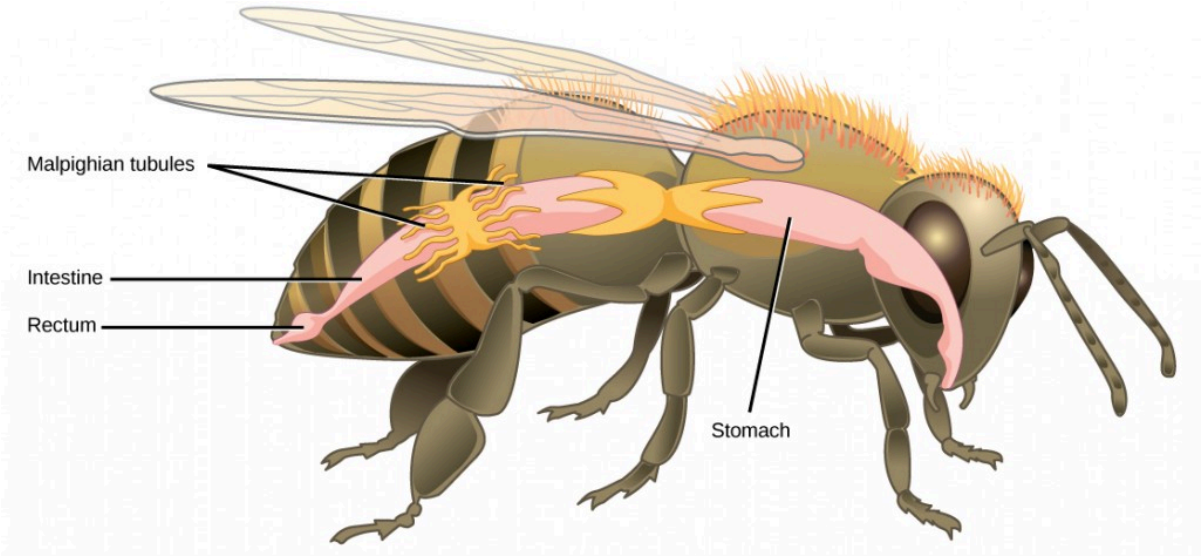


Figure 22.11. Malpighian tubules of insects and other terrestrial arthropods remove nitrogenous wastes and other solutes from the hemolymph. Na+ and/or K+ ions are actively transported into the lumen of the tubules. Water then enters the tubules via osmosis, forming urine. The urine passes through the intestine, and into the rectum. There, nutrients diffuse back into the hemolymph. Na+ and/or K+ ions are pumped into the hemolymph, and water follows. The concentrated waste is then excreted.

Concept in Action

Visit this site to see a dissected cockroach, including a close-up look at its Malpighian tubules.

Summary

Many systems have evolved for excreting wastes that are simpler than the kidney and urinary systems of vertebrate animals. The simplest system is that of contractile vacuoles present in microorganisms. Flame cells and nephridia in worms perform excretory functions and maintain osmotic balance. Some insects have evolved Malpighian tubules to excrete wastes and maintain osmotic balance.

Exercises

1. Active transport of K^+ in Malpighian tubules ensures that:
    1. water follows K^+ to make urine
    2. osmotic balance is maintained between waste matter and bodily fluids
    3. both a and b
    4. neither a nor b
2. Contractile vacuoles in microorganisms:
    1. exclusively perform an excretory function
    2. can perform many functions, one of which is excretion of metabolic wastes
    3. originate from the cell membrane
    4. both b and c
3. Flame cells are primitive excretory organs found in _____.
    1. arthropods
    2. annelids
    3. mammals
    4. flatworms
4. Why might specialized organs have evolved for excretion of wastes?
5. Explain two different excretory systems other than the kidneys.

Answers

1. C
2. D
3. D
4. The removal of wastes, which could otherwise be toxic to an organism, is extremely important for survival. Having organs that specialize in this process and that operate separately from other organs provides a measure of safety for the organism.
5. (1) Microorganisms engulf food by endocytosis—the formation of vacuoles by involution of the cell membrane within the cells. The same vacuoles interact and exchange metabolites with the intracellular environment. Cellular wastes are excreted by exocytosis when the vacuoles merge with the cell membrane and excrete wastes into the environment. (2) Flatworms have an excretory system that consists of two tubules. The cells in the tubules are called flame cells; they have a cluster of cilia that propel waste matter down the tubules and out of the body. (3) Annelids have nephridia which have a tubule with cilia. Excretion occurs through a pore called the nephridiopore. Annelids have a system for tubular reabsorption by a capillary network before excretion. (4) Malpighian tubules are found in some species of arthropods. They are usually found in pairs, and the number of tubules varies with the species of insect. Malpighian tubules are convoluted, which increases their surface area, and they are lined with microvilli for reabsorption and maintenance of osmotic balance. Metabolic wastes like uric acid freely diffuse into the tubules. Potassium ion pumps line the tubules, which actively transport out K^+ ions, and water follows to form urine. Water and electrolytes are reabsorbed when these organisms are faced with low-water environments, and uric acid is excreted as a thick paste or powder. By not dissolving wastes in water, these organisms conserve water.

Glossary

flame cell
(also, protonephridia) excretory cell found in flatworms

Malpighian tubule
excretory tubules found in arthropods

microvilli
cellular processes that increase the surface area of cells

nephridia
excretory structures found in annelids

nephridiopore
pore found at the end of nephridia

22.4. Nitrogenous Wastes

Learning Objectives

By the end of this section, you will be able to:

- Compare and contrast the way in which aquatic animals and terrestrial animals can eliminate toxic ammonia from their systems
- Compare the major byproduct of ammonia metabolism in vertebrate animals to that of birds, insects, and reptiles

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. Animals that live in aquatic environments tend to release ammonia into the water. Animals that excrete ammonia are said to be **ammonotelic**. Terrestrial organisms have evolved other mechanisms to excrete nitrogenous wastes. The animals must 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: The Urea Cycle

The **urea cycle**

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 → $H_2N\text{-}CO\text{-}NH_2$ (urea) + 2 ADP + 4 P_i + AMP.

The urea cycle utilizes five intermediate steps, catalyzed by five different enzymes, to convert ammonia to urea, as shown in Figure 22.12. The amino acid L-ornithine gets converted into different intermediates before being regenerated at the end of the urea cycle. Hence, the urea cycle is also referred to as the ornithine cycle. The enzyme ornithine transcarbamylase catalyzes a key step in the urea cycle and its deficiency can lead to accumulation of toxic levels of ammonia in the body. The first two reactions occur in the mitochondria and the last three reactions occur in the cytosol. Urea concentration in the blood, called **blood urea nitrogen** or BUN, is used as an indicator of kidney function.

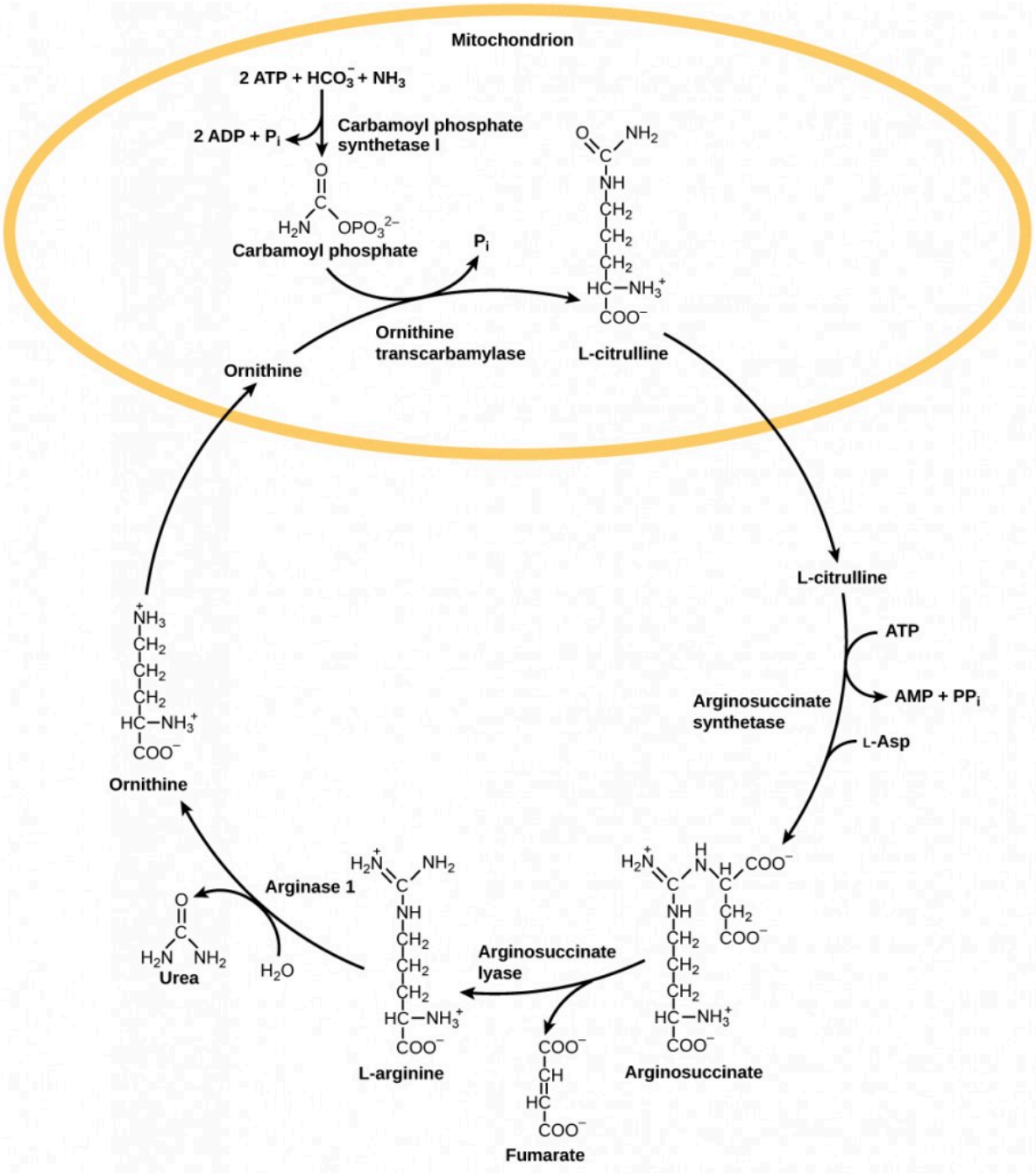


Figure 22.12. The urea cycle converts ammonia to urea.

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 22.13.

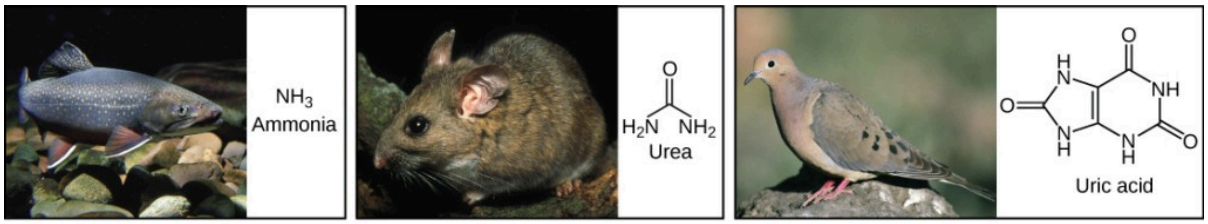


Figure 22.13. Nitrogenous waste is excreted in different forms by different species. These include (a) ammonia, (b) urea, and (c) uric acid. (credit a: modification of work by Eric Engbretson, USFWS; credit b: modification of work by B. "Moose" Peterson, USFWS; credit c: modification of work by Dave Menke, USFWS)

Gout

Mammals use uric acid crystals as an **antioxidant** in their cells. However, too much uric acid tends to form kidney stones and may also cause a painful condition called gout, where uric acid crystals accumulate in the joints, as illustrated in Figure 22.14. Food choices that reduce the amount of nitrogenous bases 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.

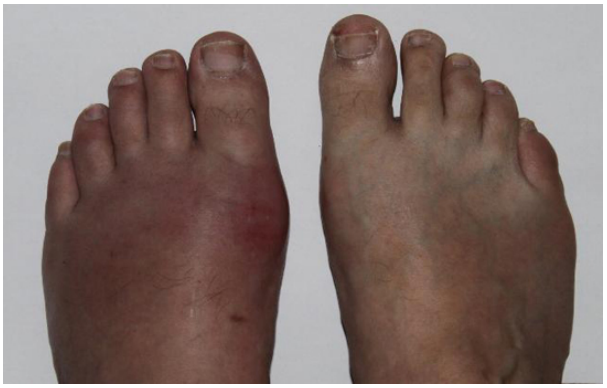


Figure 22.14. Gout causes the inflammation visible in this person's left big toe joint. (credit: "Gonzosft"/Wikimedia Commons)

Summary

Ammonia is the waste produced by metabolism of nitrogen-containing compounds like proteins and nucleic acids. While aquatic animals can easily excrete ammonia into their watery surroundings, terrestrial animals have evolved special mechanisms to eliminate the toxic ammonia from their systems. Urea is the major byproduct of ammonia metabolism in vertebrate animals. Uric acid is the major byproduct of ammonia metabolism in birds, terrestrial arthropods, and reptiles.

Exercises

1. BUN is _____.
    1. blood urea nitrogen
    2. blood uric acid nitrogen
    3. an indicator of blood volume
    4. an indicator of blood pressure
2. Human beings accumulate _____ before excreting nitrogenous waste.
    1. nitrogen
    2. ammonia
    3. urea
    4. uric acid
3. In terms of evolution, why might the urea cycle have evolved in organisms?
4. Compare and contrast the formation of urea and uric acid.

Answers

1. A
2. C
3. It is believed that 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.
4. The urea cycle is the primary mechanism by which mammals convert ammonia to urea. Urea is made in the liver and excreted in urine. The urea cycle utilizes five intermediate steps, catalyzed by five different enzymes, to convert ammonia to urea. Birds, reptiles, and insects, on the other hand, convert toxic ammonia to uric acid instead of urea. Conversion of ammonia to uric acid requires more energy and is much more complex than conversion of ammonia to urea.

Glossary

ammonia
compound made of one nitrogen atom and three hydrogen atoms

ammonotelic
describes an animal that excretes ammonia as the primary waste material

antioxidant
agent that prevents cell destruction by reactive oxygen species

blood urea nitrogen (BUN)
estimate of urea in the blood and an indicator of kidney function

urea cycle
pathway by which ammonia is converted to urea

ureotelic
describes animals that secrete urea as the primary nitrogenous waste material

uric acid
byproduct of ammonia metabolism in birds, insects, and reptiles

22.5. Hormonal Control of Osmoregulatory Functions

Learning Objectives

By the end of this section, you will be able to:

- Explain how hormonal cues help the kidneys synchronize the osmotic needs of the body
- Describe how hormones like epinephrine, norepinephrine, renin-angiotensin, aldosterone, anti-diuretic hormone, and atrial natriuretic peptide help regulate waste elimination, maintain correct osmolarity, and perform other osmoregulatory functions

While the kidneys operate to maintain osmotic balance and blood pressure in the body, they also act in concert with hormones. Hormones are small molecules that act as messengers within the body. Hormones are typically secreted from one cell and travel in the bloodstream to affect a target cell in another portion of the body. Different regions of the nephron bear specialized cells that have receptors to respond to chemical messengers and hormones. Table 22.1 summarizes the hormones that control the osmoregulatory functions.

Table 22.1.Hormones That Affect Osmoregulation

Hormone	Where produced	Function
Epinephrine and Norepinephrine	Adrenal medulla	Can decrease kidney function temporarily by vasoconstriction
Renin	Kidney nephrons	Increases blood pressure by acting on angiotensinogen
Angiotensin	Liver	Angiotensin II affects multiple processes and increases blood pressure
Aldosterone	Adrenal cortex	Prevents loss of sodium and water
Anti-diuretic hormone (vasopressin)	Hypothalamus (stored in the posterior pituitary)	Prevents water loss
Atrial natriuretic peptide	Heart atrium	Decreases blood pressure by acting as a vasodilator and increasing glomerular filtration rate; decreases sodium reabsorption in kidneys

Epinephrine and Norepinephrine

Epinephrine and norepinephrine are released by the adrenal medulla and nervous system respectively. They are the flight/fight hormones that are released when the body is under extreme stress. During

stress, much of the body's energy is used to combat imminent danger. Kidney function is halted temporarily by epinephrine and norepinephrine. These hormones function by acting directly on the smooth muscles of blood vessels to constrict them. Once the afferent arterioles are constricted, blood flow into the nephrons stops. These hormones go one step further and trigger the **renin-angiotensin-aldosterone** system.

Renin-Angiotensin-Aldosterone

The renin-angiotensin-aldosterone system, illustrated in Figure 22.15 proceeds through several steps to produce **angiotensin II**, which acts to stabilize blood pressure and volume. Renin (secreted by a part of the juxtaglomerular complex) is produced by the granular cells of the afferent and efferent arterioles. Thus, the kidneys control blood pressure and volume directly. Renin acts on angiotensinogen, which is made in the liver and converts it to **angiotensin I**. **Angiotensin converting enzyme (ACE)** converts angiotensin I to angiotensin II. Angiotensin II raises blood pressure by constricting blood vessels. It also triggers the release of the mineralocorticoid aldosterone from the adrenal cortex, which in turn stimulates the renal tubules to reabsorb more sodium. Angiotensin II also triggers the release of **anti-diuretic hormone (ADH)** from the hypothalamus, leading to water retention in the kidneys. It acts directly on the nephrons and decreases glomerular filtration rate. Medically, blood pressure can be controlled by drugs that inhibit ACE (called ACE inhibitors).

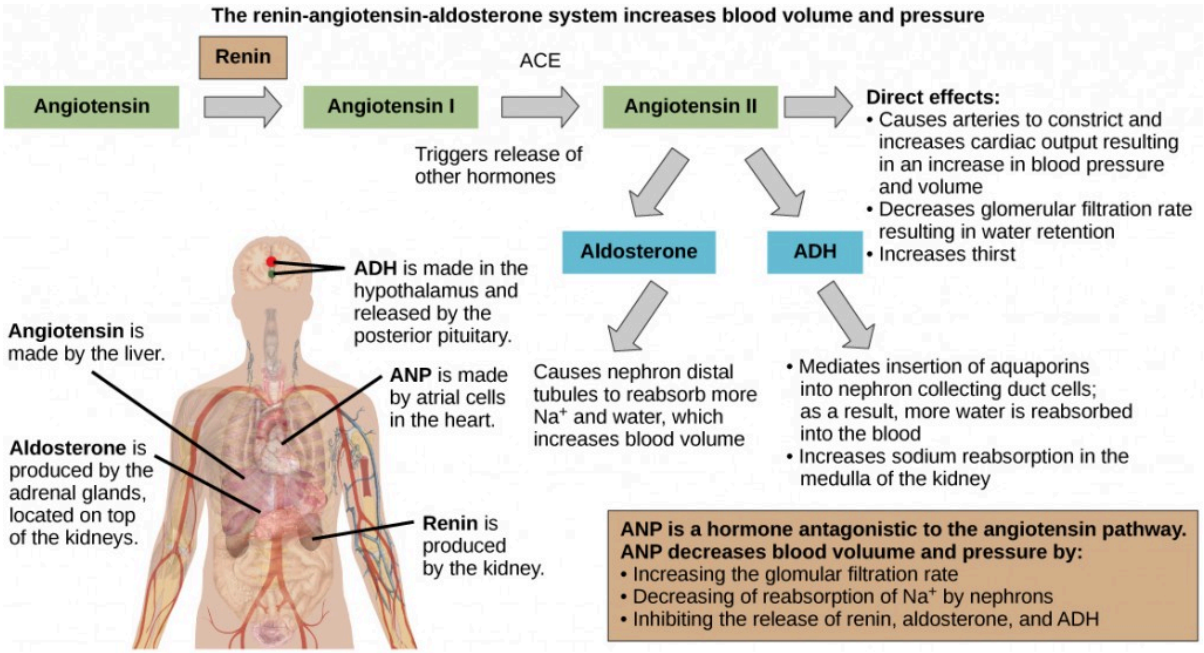


Figure 22.15. The renin-angiotensin-aldosterone system increases blood pressure and volume. The hormone ANP has antagonistic effects. (credit: modification of work by Mikael Häggström)

Mineralocorticoids

Mineralocorticoids are hormones synthesized by the adrenal cortex that affect osmotic balance. Aldosterone is a mineralocorticoid that regulates sodium levels in the blood. Almost all of the sodium in the blood is reclaimed by the renal tubules under the influence of aldosterone. Because sodium is always reabsorbed by active transport and water follows sodium to maintain osmotic balance,

aldosterone manages not only sodium levels but also the water levels in body fluids. In contrast, the aldosterone also stimulates potassium secretion concurrently with sodium reabsorption. In contrast, absence of aldosterone means that no sodium gets reabsorbed in the renal tubules and all of it gets excreted in the urine. In addition, the daily dietary potassium load is not secreted and the retention of K^+ can cause a dangerous increase in plasma K^+ concentration. Patients who have Addison's disease have a failing adrenal cortex and cannot produce aldosterone. They lose sodium in their urine constantly, and if the supply is not replenished, the consequences can be fatal.

Antidiurectic Hormone

As previously discussed, antidiuretic hormone or ADH (also called **vasopressin**), as the name suggests, helps the body conserve water when body fluid volume, especially that of blood, is low. It is formed by the hypothalamus and is stored and released from the posterior pituitary. It acts by inserting aquaporins in the collecting ducts and promotes reabsorption of water. ADH also acts as a vasoconstrictor and increases blood pressure during hemorrhaging.

Atrial Natriuretic Peptide Hormone

The atrial natriuretic peptide (ANP) lowers blood pressure by acting as a **vasodilator**. It is released by cells in the atrium of the heart in response to high blood pressure and in patients with sleep apnea. ANP affects salt release, and because water passively follows salt to maintain osmotic balance, it also has a diuretic effect. ANP also prevents sodium reabsorption by the renal tubules, decreasing water reabsorption (thus acting as a diuretic) and lowering blood pressure. Its actions suppress the actions of aldosterone, ADH, and renin.

Summary

Hormonal cues help the kidneys synchronize the osmotic needs of the body. Hormones like epinephrine, norepinephrine, renin-angiotensin, aldosterone, anti-diuretic hormone, and atrial natriuretic peptide help regulate the needs of the body as well as the communication between the different organ systems.

Exercises

1. Renin is made by _____.
    1. granular cells of the juxtaglomerular apparatus
    2. the kidneys
    3. the nephrons
    4. All of the above.
2. Patients with Addison's disease _____.
    1. retain water

2. retain salts
3. lose salts and water
4. have too much aldosterone

3. Which hormone elicits the “fight or flight” response?

1. epinephrine
2. mineralcorticoids
3. anti-diuretic hormone
4. thyroxine

4. Describe how hormones regulate blood pressure, blood volume, and kidney function.
5. How does the renin-angiotensin-aldosterone mechanism function? Why is it controlled by the kidneys?

Answers

1. A
2. C
3. A
4. Hormones are small molecules that act as messengers within the body. Different regions of the nephron bear specialized cells, which have receptors to respond to chemical messengers and hormones. The hormones carry messages to the kidney. These hormonal cues help the kidneys synchronize the osmotic needs of the body. Hormones like epinephrine, norepinephrine, renin-angiotensin, aldosterone, anti-diuretic hormone, and atrial natriuretic peptide help regulate the needs of the body as well as the communication between the different organ systems.
5. The renin-angiotensin-aldosterone system acts through several steps to produce angiotensin II, which acts to stabilize blood pressure and volume. Thus, the kidneys control blood pressure and volume directly. Renin acts on angiotensinogen, which is made in the liver and converts it to angiotensin I. ACE (angiotensin converting enzyme) converts angiotensin I to angiotensin II. Angiotensin II raises blood pressure by constricting blood vessels. It triggers the release of aldosterone from the adrenal cortex, which in turn stimulates the renal tubules to reabsorb more sodium. Angiotensin II also triggers the release of anti-diuretic hormone from the hypothalamus, which leads to water retention. It acts directly on the nephrons and decreases GFR.

Glossary

angiotensin II
molecule that affects different organs to increase blood pressure

angiotensin I
product in the renin-angiotensin-aldosterone pathway

angiotensin converting enzyme (ACE)
enzyme that converts angiotensin I to angiotensin II

anti-diuretic hormone (ADH)
hormone that prevents the loss of water

renin-angiotensin-aldosterone
biochemical pathway that activates angiotensin II, which increases blood pressure

vasodilator
compound that increases the diameter of blood vessels

vasopressin
another name for anti-diuretic hormone

Chapter 23. The Immune System

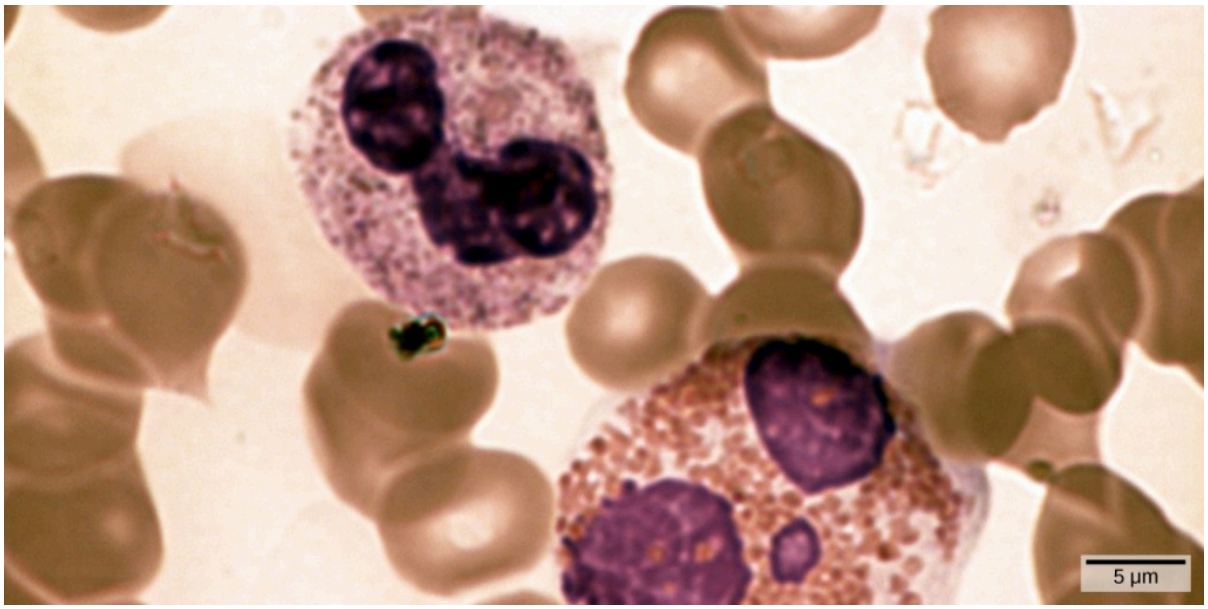


Figure 23.1. In this compound light micrograph purple-stained neutrophil (upper left) and eosinophil (lower right) are white blood cells that float among red blood cells in this blood smear. Neutrophils provide an early, rapid, and nonspecific defense against invading pathogens. Eosinophils play a variety of roles in the immune response. Red blood cells are about 7–8 µm in diameter, and a neutrophil is about 10–12µm. (credit: modification of work by Dr. David Csaba)

Introduction

The environment consists of numerous **pathogens**, which are agents, usually microorganisms, that cause diseases in their hosts. A **host** is the organism that is invaded and often harmed by a pathogen. Pathogens include bacteria, protists, fungi and other infectious organisms. We are constantly exposed to pathogens in food and water, on surfaces, and in the air. Mammalian immune systems evolved for protection from such pathogens; they are composed of an extremely diverse array of specialized cells and soluble molecules that coordinate a rapid and flexible defense system capable of providing protection from a majority of these disease agents.

Components of the immune system constantly search the body for signs of pathogens. When pathogens are found, immune factors are mobilized to the site of an infection. The immune factors identify the nature of the pathogen, strengthen the corresponding cells and molecules to combat it efficiently, and then halt the immune response after the infection is cleared to avoid unnecessary host cell damage. The immune system can remember pathogens to which it has been exposed to create a more efficient response upon re-exposure. This memory can last several decades. Features of the immune system, such as pathogen identification, specific response, amplification, retreat, and remembrance are essential for survival against pathogens. The immune response can be classified as either innate or active. The innate immune response is always present and attempts to defend against all pathogens rather than focusing on specific ones. Conversely, the adaptive immune response stores information about past infections and mounts pathogen-specific defenses.

Glossary

host
an organism that is invaded by a pathogen or parasite

pathogen
an agent, usually a microorganism, that causes disease in the organisms that they invade

23.1. Innate Immune Response

Learning Objectives

By the end of this section, you will be able to:

- Describe physical and chemical immune barriers
- Explain immediate and induced innate immune responses
- Discuss natural killer cells
- Describe major histocompatibility class I molecules
- Summarize how the proteins in a complement system function to destroy extracellular pathogens

The immune system comprises both innate and adaptive immune responses. **Innate immunity** occurs naturally because of genetic factors or physiology; it is not induced by infection or vaccination but works to reduce the workload for the adaptive immune response. Both the innate and adaptive levels of the immune response involve secreted proteins, receptor-mediated signaling, and intricate cell-to-cell communication. The innate immune system developed early in animal evolution, roughly a billion years ago, as an essential response to infection. Innate immunity has a limited number of specific targets: any pathogenic threat triggers a consistent sequence of events that can identify the type of pathogen and either clear the infection independently or mobilize a highly specialized adaptive immune response. For example, tears and mucus secretions contain microbicidal factors.

Physical and Chemical Barriers

Before any immune factors are triggered, the skin functions as a continuous, impassable barrier to potentially infectious pathogens. Pathogens are killed or inactivated on the skin by desiccation (drying out) and by the skin's acidity. In addition, beneficial microorganisms that coexist on the skin compete with invading pathogens, preventing infection. Regions of the body that are not protected by skin (such as the eyes and mucus membranes) have alternative methods of defense, such as tears and mucus secretions that trap and rinse away pathogens, and cilia in the nasal passages and respiratory tract that push the mucus with the pathogens out of the body. Throughout the body are other defenses, such as the low pH of the stomach (which inhibits the growth of pathogens), blood proteins that bind and disrupt bacterial cell membranes, and the process of urination (which flushes pathogens from the urinary tract).

Despite these barriers, pathogens may enter the body through skin abrasions or punctures, or by collecting on mucosal surfaces in large numbers that overcome the mucus or cilia. Some pathogens have evolved specific mechanisms that allow them to overcome physical and chemical barriers. When pathogens do enter the body, the innate immune system responds with inflammation, pathogen engulfment, and secretion of immune factors and proteins.

Pathogen Recognition

An infection may be intracellular or extracellular, depending on the pathogen. All viruses infect cells and replicate within those cells (intracellularly), whereas bacteria and other parasites may replicate intracellularly or extracellularly, depending on the species. The innate immune system must respond accordingly: by identifying the extracellular pathogen and/or by identifying host cells that have already been infected. When a pathogen enters the body, cells in the blood and lymph detect the specific **pathogen-associated molecular patterns (PAMPs)** on the pathogen's surface. PAMPs are carbohydrate, polypeptide, and nucleic acid "signatures" that are expressed by viruses, bacteria, and parasites but which differ from molecules on host cells. The immune system has specific cells, described in Figure 23.2 and shown in Figure 23.3, with receptors that recognize these PAMPs. A **macrophage** is a large phagocytic cell that engulfs foreign particles and pathogens. Macrophages recognize PAMPs via complementary **pattern recognition receptors (PRRs)**. PRRs are molecules on macrophages and dendritic cells which are in contact with the external environment. A **monocyte** is a type of white blood cell that circulates in the blood and lymph and differentiates into macrophages after it moves into infected tissue. Dendritic cells bind molecular signatures of pathogens and promote pathogen engulfment and destruction. Toll-like receptors (TLRs) are a type of PRR that recognizes molecules that are shared by pathogens but distinguishable from host molecules). TLRs are present in invertebrates as well as vertebrates, and appear to be one of the most ancient components of the immune system. TLRs have also been identified in the mammalian nervous system.

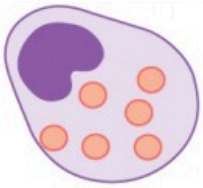
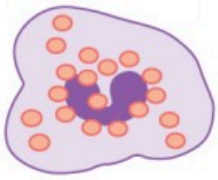
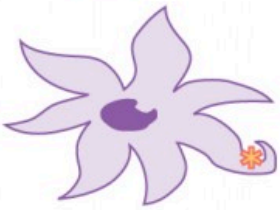
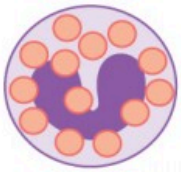
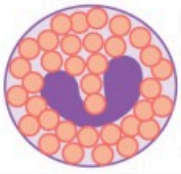
Cell type	Characteristics	Location	Image
Mast cell	Dilates blood vessels and induces inflammation through release of histamines and heparin. Recruits macrophages and neutrophils. Involved in wound healing and defense against pathogens but can also be responsible for allergic reactions.	Connective tissues, mucous membranes	
Macrophage	Phagocytic cell that consumes foreign pathogens and cancer cells. Stimulates response of other immune cells.	Migrates from blood vessels into tissues.	
Natural killer cell	Kills tumor cells and virus-infected cells.	Circulates in blood and migrates into tissues.	
Dendritic cell	Presents antigens on its surface, thereby triggering adaptive immunity.	Present in epithelial tissue, including skin, lung and tissues of the digestive tract. Migrates to lymph nodes upon activation.	
Monocyte	Differentiates into macrophages and dendritic cells in response to inflammation.	Stored in spleen, moves through blood vessels to infected tissues.	
Neutrophil	First responders at the site of infection or trauma, this abundant phagocytic cell represents 50-60 percent of all leukocytes. Releases toxins that kill or inhibit bacteria and fungi and recruits other immune cells to the site of infection.	Migrates from blood vessels into tissues.	
Basophil	Responsible for defense against parasites. Releases histamines that cause inflammation and may be responsible for allergic reactions.	Circulates in blood and migrates to tissues.	
Eosinophil	Releases toxins that kill bacteria and parasites but also causes tissue damage.	Circulates in blood and migrates to tissues.	

Figure 23.2. The characteristics and location of cells involved in the innate immune system are described. (credit: modification of work by NIH)

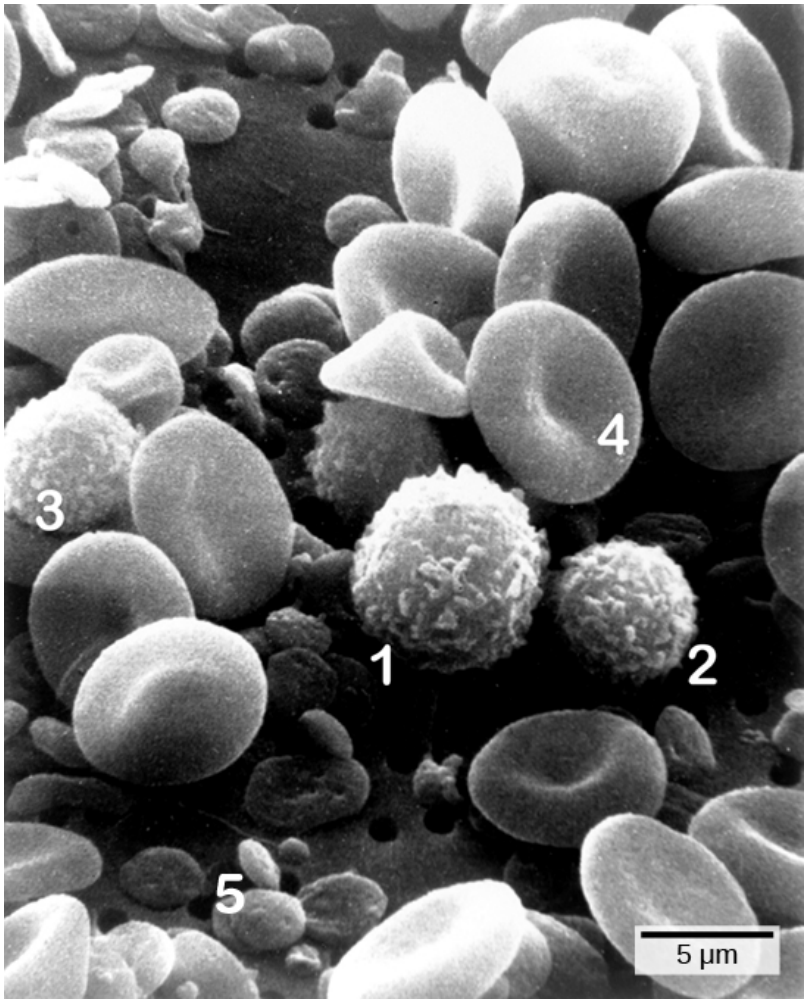


Figure 23.3. Cells of the blood include (1) monocytes, (2) lymphocytes, (3) neutrophils, (4) red blood cells, and (5) platelets. Note the very similar morphologies of the leukocytes (1, 2, 3). (credit: modification of work by Bruce Wetzel, Harry Schaefer, NCI; scale-bar data from Matt Russell)

Cytokine Release Affect

The binding of PRRs with PAMPs triggers the release of cytokines, which signal that a pathogen is present and needs to be destroyed along with any infected cells. A **cytokine** is a chemical messenger that regulates cell differentiation (form and function), proliferation (production), and gene expression to affect immune responses. At least 40 types of cytokines exist in humans that differ in terms of the cell type that produces them, the cell type that responds to them, and the changes they produce. One type cytokine, interferon, is illustrated in Figure 23.4.

One subclass of cytokines is the interleukin (IL), so named because they mediate interactions between leukocytes (white blood cells). Interleukins are involved in bridging the innate and adaptive immune responses. In addition to being released from cells after PAMP recognition, cytokines are released by the infected cells which bind to nearby uninfected cells and induce those cells to release cytokines, which results in a cytokine burst.

A second class of early-acting cytokines is interferons, which are released by infected cells as a warning to nearby uninfected cells. One of the functions of an **interferon** is to inhibit viral replication. They also have other important functions, such as tumor surveillance. Interferons work by signaling neighboring uninfected cells to destroy RNA and reduce protein synthesis, signaling neighboring infected cells to undergo apoptosis (programmed cell death), and activating immune cells.

In response to interferons, uninfected cells alter their gene expression, which increases the cells' resistance to infection. One effect of interferon-induced gene expression is a sharply reduced cellular protein synthesis. Virally infected cells produce more viruses by synthesizing large quantities of viral proteins. Thus, by reducing protein synthesis, a cell becomes resistant to viral infection.

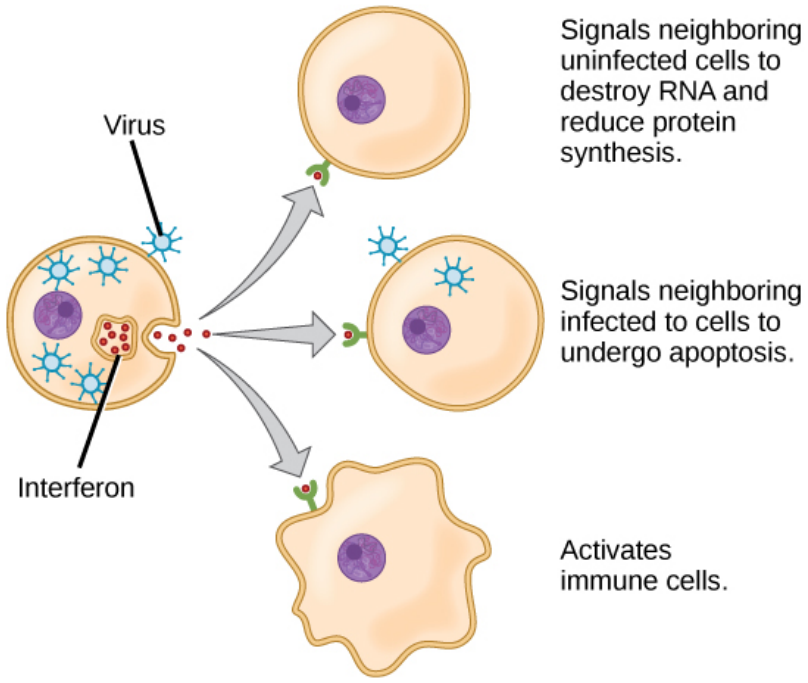


Figure 23.4. Interferons are cytokines that are released by a cell infected with a virus. Response of neighboring cells to interferon helps stem the infection.

Phagocytosis and Inflammation

The first cytokines to be produced are pro-inflammatory; that is, they encourage **inflammation**, the localized redness, swelling, heat, and pain that result from the movement of leukocytes and fluid through increasingly permeable capillaries to a site of infection. The population of leukocytes that arrives at an infection site depends on the nature of the infecting pathogen. Both macrophages and dendritic cells engulf pathogens and cellular debris through phagocytosis. A **neutrophil** is also a phagocytic leukocyte that engulfs and digests pathogens. Neutrophils, shown in Figure 23.3, are the most abundant leukocytes of the immune system. Neutrophils have a nucleus with two to five lobes, and they contain organelles, called lysosomes, that digest engulfed pathogens. An **eosinophil** is a leukocyte that works with other eosinophils to surround a parasite; it is involved in the allergic response and in protection against helminthes (parasitic worms).

Neutrophils and eosinophils are particularly important leukocytes that engulf large pathogens, such as

bacteria and fungi. A **mast cell** is a leukocyte that produces inflammatory molecules, such as histamine, in response to large pathogens. A **basophil** is a leukocyte that, like a neutrophil, releases chemicals to stimulate the inflammatory response as illustrated in Figure 23.5. Basophils are also involved in allergy and hypersensitivity responses and induce specific types of inflammatory responses. Eosinophils and basophils produce additional inflammatory mediators to recruit more leukocytes. A hypersensitive immune response to harmless antigens, such as in pollen, often involves the release of histamine by basophils and mast cells.

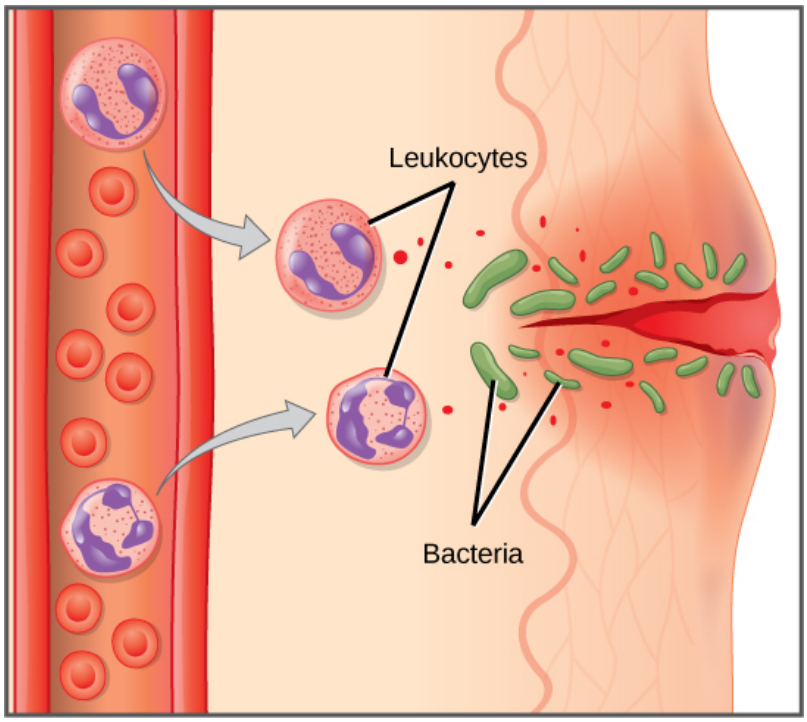


Figure 23.5. In response to a cut, mast cells secrete histamines that cause nearby capillaries to dilate. Neutrophils and monocytes leave the capillaries. Monocytes mature into macrophages. Neutrophils, dendritic cells and macrophages release chemicals to stimulate the inflammatory response. Neutrophils and macrophages also consume invading bacteria by phagocytosis.

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. These effects may have evolved because the symptoms encourage the individual to rest and prevent them from spreading the infection to others. Cytokines also increase the core body temperature, causing a fever, which causes the liver to withhold iron from the blood. Without iron, certain pathogens, such as some bacteria, are unable to replicate; this is called nutritional immunity.

Concept in Action

Watch this 23-second stop-motion video showing a neutrophil that searches for and engulfs fungus spores during an elapsed time of about 79 minutes.

Natural Killer Cells

Lymphocytes are leukocytes that are histologically identifiable by their large, darkly staining nuclei; they are small cells with very little cytoplasm, as shown in Figure 23.6. Infected cells are identified and destroyed by **natural killer (NK) cells**, lymphocytes that can kill cells infected with viruses or tumor cells (abnormal cells that uncontrollably divide and invade other tissue). T cells and B cells of the adaptive immune system also are classified as lymphocytes. **T cells** are lymphocytes that mature in the thymus gland, and **B cells** are lymphocytes that mature in the bone marrow. NK cells identify intracellular infections, especially from viruses, by the altered expression of **major histocompatibility class (MHC) I molecules** on the surface of infected cells. MHC I molecules are proteins on the surfaces of all nucleated cells, thus they are scarce on red blood cells and platelets which are non-nucleated. The function of MHC I molecules is to display fragments of proteins from the infectious agents within the cell to T-cells; healthy cells will be ignored, while “non-self” or foreign proteins will be attacked by the immune system. MHC II molecules are found mainly on cells containing antigens (“non-self proteins”) and on lymphocytes. **MHC II molecules** interact with helper T-cells to trigger the appropriate immune response, which may include the inflammatory response.

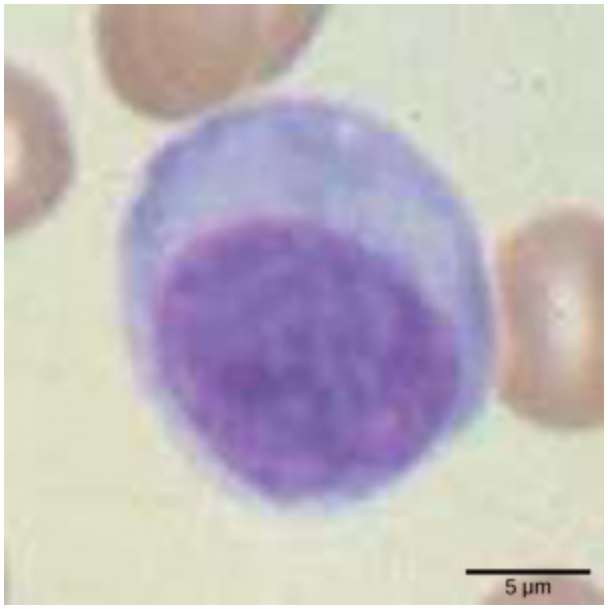


Figure 23.6. Lymphocytes, such as NK cells, are characterized by their large nuclei that actively absorb Wright stain and therefore appear dark colored under a microscope.

An infected cell (or a tumor cell) is usually incapable of synthesizing and displaying MHC I molecules appropriately. The metabolic resources of cells infected by some viruses produce proteins that interfere with MHC I processing and/or trafficking to the cell surface. The reduced MHC I on host cells varies from virus to virus and results from active inhibitors being produced by the viruses. This process can deplete host MHC I molecules on the cell surface, which NK cells detect as “unhealthy” or “abnormal” while searching for cellular MHC I molecules. Similarly, the dramatically altered gene expression of tumor cells leads to expression of extremely deformed or absent MHC I molecules that also signal “unhealthy” or “abnormal.”

NK cells are always active; an interaction with normal, intact MHC I molecules on a healthy cell disables the killing sequence, and the NK cell moves on. After the NK cell detects an infected or tumor cell, its cytoplasm secretes granules comprised of **perforin**, a destructive protein that creates a pore in the target cell. Granzymes are released along with the perforin in the immunological synapse. A **granzyme** is a protease that digests cellular proteins and induces the target cell to undergo programmed cell death, or apoptosis. Phagocytic cells then digest the cell debris left behind. NK cells are constantly patrolling the body and are an effective mechanism for controlling potential infections and preventing cancer progression.

Complement

An array of approximately 20 types of soluble proteins, called a **complement system**, functions to destroy extracellular pathogens. Cells of the liver and macrophages synthesize 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 antibody response of the adaptive immune system. Complement proteins bind to the surfaces of microorganisms and are particularly attracted to pathogens that are already bound by

antibodies. Binding of complement proteins occurs in a specific and highly regulated sequence, with each successive protein being activated by cleavage and/or structural changes induced upon binding of the preceding protein(s). After the first few complement proteins bind, a cascade of sequential binding events follows in which the pathogen rapidly becomes coated in complement proteins.

Complement proteins perform several functions. The proteins serve as a marker to indicate the presence of a pathogen to phagocytic cells, such as macrophages and B cells, and enhance engulfment; this process is called **opsonization**. Certain complement proteins can combine to form attack complexes that open pores in microbial cell membranes. These structures destroy pathogens by causing their contents to leak, as illustrated in Figure 23.7.

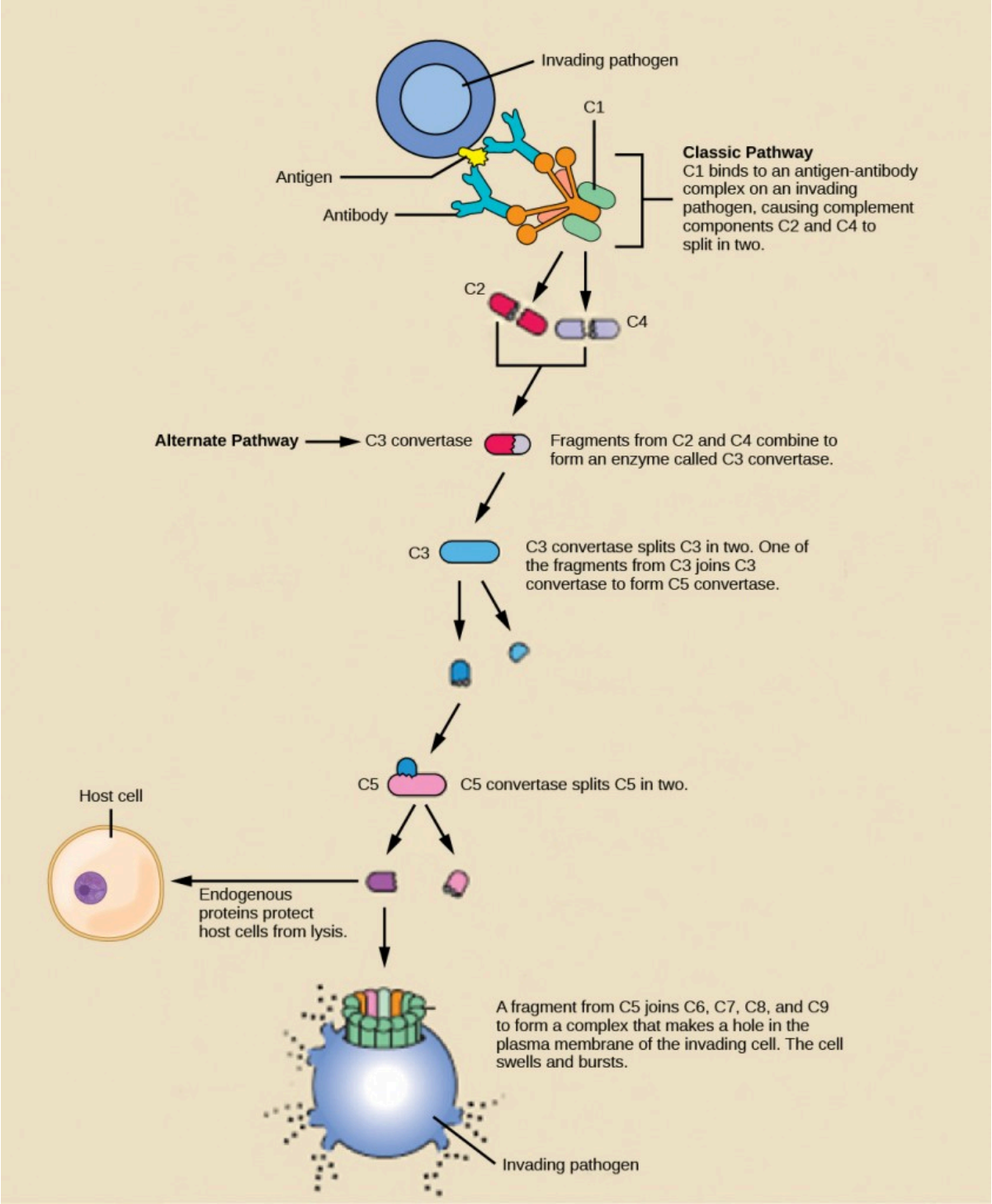


Figure 23.7. The classic pathway for the complement cascade involves the attachment of several initial complement proteins to an antibody-bound pathogen followed by rapid activation and binding of many more complement proteins and the creation of destructive pores in the microbial cell envelope and cell wall. The alternate pathway does not involve antibody activation. Rather, C3 convertase spontaneously breaks down C3. Endogenous regulatory proteins prevent the complement complex from binding to host cells. Pathogens lacking these regulatory proteins are lysed. (credit: modification of work by NIH)

Summary

The innate immune system serves as a first responder to pathogenic threats that bypass natural physical and chemical barriers of the body. Using a combination of cellular and molecular attacks, the innate immune system identifies the nature of a pathogen and responds with inflammation, phagocytosis, cytokine release, destruction by NK cells, and/or a complement system. When innate mechanisms are insufficient to clear an infection, the adaptive immune response is informed and mobilized.

Exercises

1. Which of the following is a barrier against pathogens provided by the skin?
    1. high pH
    2. mucus
    3. tears
    4. desiccation
2. Although interferons have several effects, they are particularly useful against infections with which type of pathogen?
    1. bacteria
    2. viruses
    3. fungi
    4. helminths
3. Which organelle do phagocytes use to digest engulfed particles?
    1. lysosome
    2. nucleus
    3. endoplasmic reticulum
    4. mitochondria
4. Which innate immune system component uses MHC I molecules directly in its defense strategy?
    1. macrophages
    2. neutrophils
    3. NK cells
    4. interferon
5. Different MHC I molecules between donor and recipient cells can lead to rejection of a transplanted organ or tissue. Suggest a reason for this.
6. If a series of genetic mutations prevented some, but not all, of the complement proteins from binding antibodies or pathogens, would the entire complement system be compromised?

Answers

1. D
2. B
3. A
4. C
5. If the MHC I molecules expressed on donor cells differ from the MHC I molecules expressed on recipient cells, NK cells may identify the donor cells as “non-self” and produce perforin and granzymes to induce the donor cells to undergo apoptosis, which would destroy the transplanted organ.
6. The entire complement system would probably be affected even when only a few members were mutated such that they could no longer bind. Because the complement involves the binding of activated proteins in a specific sequence, when one or more proteins in the sequence are absent, the subsequent proteins would be incapable of binding to elicit the complement’s pathogen-destructive effects.

Glossary

B cell
lymphocyte that matures in the bone marrow and differentiates into antibody-secreting plasma cells

basophil
leukocyte that releases chemicals usually involved in the inflammatory response

complement system
array of approximately 20 soluble proteins of the innate immune system that enhance phagocytosis, bore holes in pathogens, and recruit lymphocytes; enhances the adaptive response when antibodies are produced

cytokine
chemical messenger that regulates cell differentiation, proliferation, gene expression, and cell trafficking to effect immune responses

eosinophil
leukocyte that responds to parasites and is involved in the allergic response

granzyme
protease that enters target cells through perforin and induces apoptosis in the target cells; used by NK cells and killer T cells

inflammation
localized redness, swelling, heat, and pain that results from the movement of leukocytes and fluid through opened capillaries to a site of infection

innate immunity
immunity that occurs naturally because of genetic factors or physiology, and is not induced by infection or vaccination

interferon
cytokine that inhibits viral replication and modulates the immune response

lymphocyte
leukocyte that is histologically identifiable by its large nuclei; it is a small cell with very little cytoplasm

macrophage
large phagocytic cell that engulfs foreign particles and pathogens

major histocompatibility class (MHC) I/II molecule
protein found on the surface of all nucleated cells (I) or specifically on antigen-presenting cells (II) that signals to immune cells whether the cell is healthy/normal or is infected/cancerous; it provides the appropriate template into which antigens can be loaded for recognition by lymphocytes

mast cell
leukocyte that produces inflammatory molecules, such as histamine, in response to large pathogens and allergens

monocyte
type of white blood cell that circulates in the blood and lymph and differentiates into macrophages after it moves into infected tissue

natural killer (NK) cell
lymphocyte that can kill cells infected with viruses or tumor cells

neutrophil
phagocytic leukocyte that engulfs and digests pathogens

opsonization
process that enhances phagocytosis using proteins to indicate the presence of a pathogen to phagocytic cells

pathogen-associated molecular pattern (PAMP)
carbohydrate, polypeptide, and nucleic acid “signature” that is expressed by viruses, bacteria, and parasites but differs from molecules on host cells

pattern recognition receptor (PRR)
molecule on macrophages and dendritic cells that binds molecular signatures of pathogens and promotes pathogen engulfment and destruction

perforin
destructive protein that creates a pore in the target cell; used by NK cells and killer T cells

T cell
lymphocyte that matures in the thymus gland; one of the main cells involved in the adaptive immune system

MHC II molecules

23.2. Adaptive Immune Response

Learning Objectives

By the end of this section, you will be able to:

- Explain adaptive immunity
- Compare and contrast adaptive and innate immunity
- Describe cell-mediated immune response and humoral immune response
- Describe immune tolerance

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 pathogens and has memory. **Adaptive immunity** is an immunity that occurs after exposure to an antigen either from a pathogen or a vaccination. 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 carried out by T cells, and the **humoral immune response**, which is controlled by activated B cells and antibodies. Activated T cells and B cells that are specific to molecular structures on the pathogen proliferate and attack the invading pathogen. Their attack can kill pathogens directly or secrete antibodies that enhance the phagocytosis of pathogens and disrupt the infection. Adaptive immunity also involves a memory to provide the host with long-term protection from reinfection with the same type of pathogen; on re-exposure, this memory will facilitate an efficient and quick response.

Antigen-presenting Cells

Unlike NK cells of the innate immune system, B cells (B lymphocytes) are a type of white blood cell that gives rise to antibodies, whereas T cells (T lymphocytes) are a type of white blood cell that plays an important role in the immune response. T cells are a key component in the cell-mediated response—the specific immune response that utilizes T cells to neutralize cells that have been infected with viruses and certain bacteria. There are three types of T cells: cytotoxic, helper, and suppressor T cells. Cytotoxic T cells destroy virus-infected cells in the cell-mediated immune response, and helper T cells play a part in activating both the antibody and the cell-mediated immune responses. Suppressor T cells deactivate T cells and B cells when needed, and thus prevent the immune response from becoming too intense.

An **antigen** is a foreign or “non-self” macromolecule that reacts with cells of the immune system. Not all antigens will provoke a response. For instance, individuals produce innumerable “self” antigens and are constantly exposed to harmless foreign antigens, such as food proteins, pollen, or dust components.

The suppression of immune responses to harmless macromolecules is highly regulated and typically prevents processes that could be damaging to the host, known as tolerance.

The innate immune system contains cells that detect potentially harmful antigens, and then inform the adaptive immune response about the presence of these antigens. An **antigen-presenting cell (APC)** is an immune cell that detects, engulfs, and informs the adaptive immune response about an infection. When a pathogen is detected, these APCs will phagocytose the pathogen and digest it to form many different fragments of the antigen. Antigen fragments will then be transported to the surface of the APC, where they will serve as an indicator to other immune cells. **Dendritic cells** are immune cells that process antigen material; they are present in the skin (Langerhans cells) and the lining of the nose, lungs, stomach, and intestines. Sometimes a dendritic cell presents on the surface of other cells to induce an immune response, thus functioning as an antigen-presenting cell. Macrophages also function as APCs. Before activation and differentiation, B cells can also function as APCs.

After phagocytosis by APCs, the phagocytic vesicle fuses with an intracellular lysosome forming phagolysosome. Within the phagolysosome, the components are broken down into fragments; the fragments are then loaded onto MHC class I or MHC class II molecules and are transported to the cell surface for antigen presentation, as illustrated in Figure 23.8. Note that T lymphocytes cannot properly respond to the antigen unless it is processed and embedded in an MHC II molecule. APCs express MHC on their surfaces, and when combined with a foreign antigen, these complexes signal a “non-self” invader. Once the fragment of antigen is embedded in the MHC II molecule, the immune cell can respond. Helper T- cells are one of the main lymphocytes that respond to antigen-presenting cells. Recall that all other nucleated cells of the body expressed MHC I molecules, which signal “healthy” or “normal.”

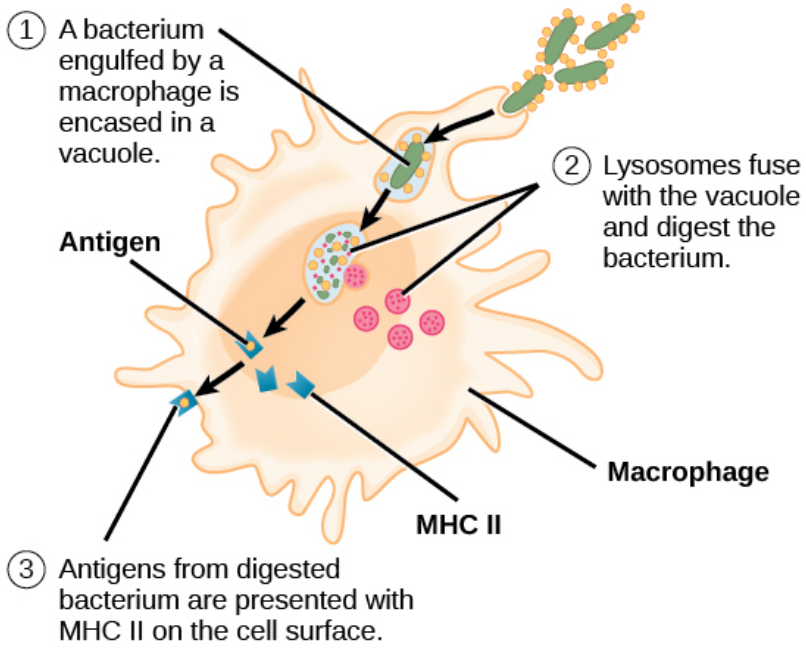


Figure 23.8. An APC, such as a macrophage, engulfs and digests a foreign bacterium. An antigen from the bacterium is presented on the cell surface in conjunction with an MHC II molecule Lymphocytes of the adaptive immune response interact with antigen-embedded MHC II molecules to mature into functional immune cells.

Concept in Action

This animation from Rockefeller University shows how dendritic cells act as sentinels in the body's immune system.

T and B Lymphocytes

Lymphocytes in human circulating blood are approximately 80 to 90 percent T cells, shown in Figure 23.9, and 10 to 20 percent B cells. Recall that the T cells are involved in the cell-mediated immune response, whereas B cells are part of the humoral immune response.

T cells encompass a heterogeneous population of cells with extremely diverse functions. Some T cells respond to APCs of the innate immune system, and indirectly induce immune responses by releasing cytokines. Other T cells stimulate B cells to prepare their own response. Another population of T cells detects APC signals and directly kills the infected cells. Other T cells are involved in suppressing inappropriate immune reactions to harmless or "self" antigens.

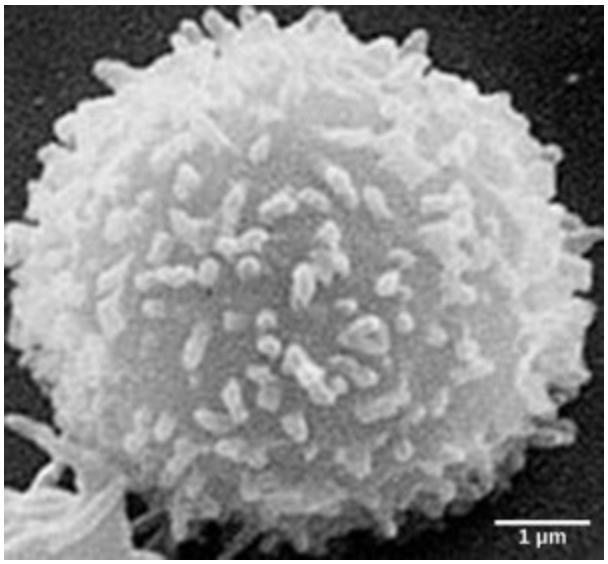


Figure 23.9. This scanning electron micrograph shows a T lymphocyte, which is responsible for the cell-mediated immune response. T cells are able to recognize antigens. (credit: modification of work by NCI; scale-bar data from Matt Russell)

T and B cells exhibit a common theme of recognition/binding of specific antigens via a complementary receptor, followed by activation and self-amplification/maturation to specifically bind to the particular antigen of the infecting pathogen. T and B lymphocytes are also similar in that each cell only expresses

one type of antigen receptor. Any individual may possess a population of T and B cells that together express a near limitless variety of antigen receptors that are capable of recognizing virtually any infecting pathogen. T and B cells are activated when they recognize small components of antigens, called **epitopes**, presented by APCs, illustrated in Figure 23.10. Note that recognition occurs at a specific epitope rather than on the entire antigen; for this reason, epitopes are known as "antigenic determinants." In the absence of information from APCs, T and B cells remain inactive, or naïve, and are unable to prepare an immune response. The requirement for information from the APCs of innate immunity to trigger B cell or T cell activation illustrates the essential nature of the innate immune response to the functioning of the entire immune system.

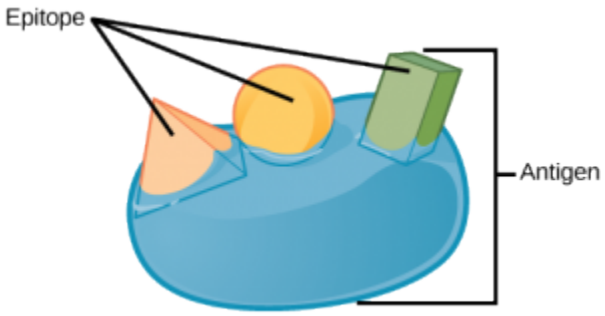


Figure 23.10. An antigen is a macromolecule that reacts with components of the immune system. A given antigen may contain several motifs that are recognized by immune cells. Each motif is an epitope. In this figure, the entire structure is an antigen, and the orange, salmon and green components projecting from it represent potential epitopes.

Naïve T cells can express one of two different molecules, CD4 or CD8, on their surface, as shown in Figure 23.11, and are accordingly classified as $CD4^+$ or $CD8^+$ cells. These molecules are important because they regulate how a T cell will interact with and respond to an APC. Naïve $CD4^+$ cells bind APCs via their antigen-embedded MHC II molecules and are stimulated to become **helper T (T_H) lymphocytes**, cells that go on to stimulate B cells (or cytotoxic T cells) directly or secrete cytokines to inform more and various target cells about the pathogenic threat. In contrast, $CD8^+$ cells engage antigen-embedded MHC I molecules on APCs and are stimulated to become **cytotoxic T lymphocytes (CTLs)**, which directly kill infected cells by apoptosis and emit cytokines to amplify the immune response. The two populations of T cells have different mechanisms of immune protection, but both bind MHC molecules via their antigen receptors called T cell receptors (TCRs). The CD4 or CD8 surface molecules differentiate whether the TCR will engage an MHC II or an MHC I molecule. Because they assist in binding specificity, the CD4 and CD8 molecules are described as coreceptors.

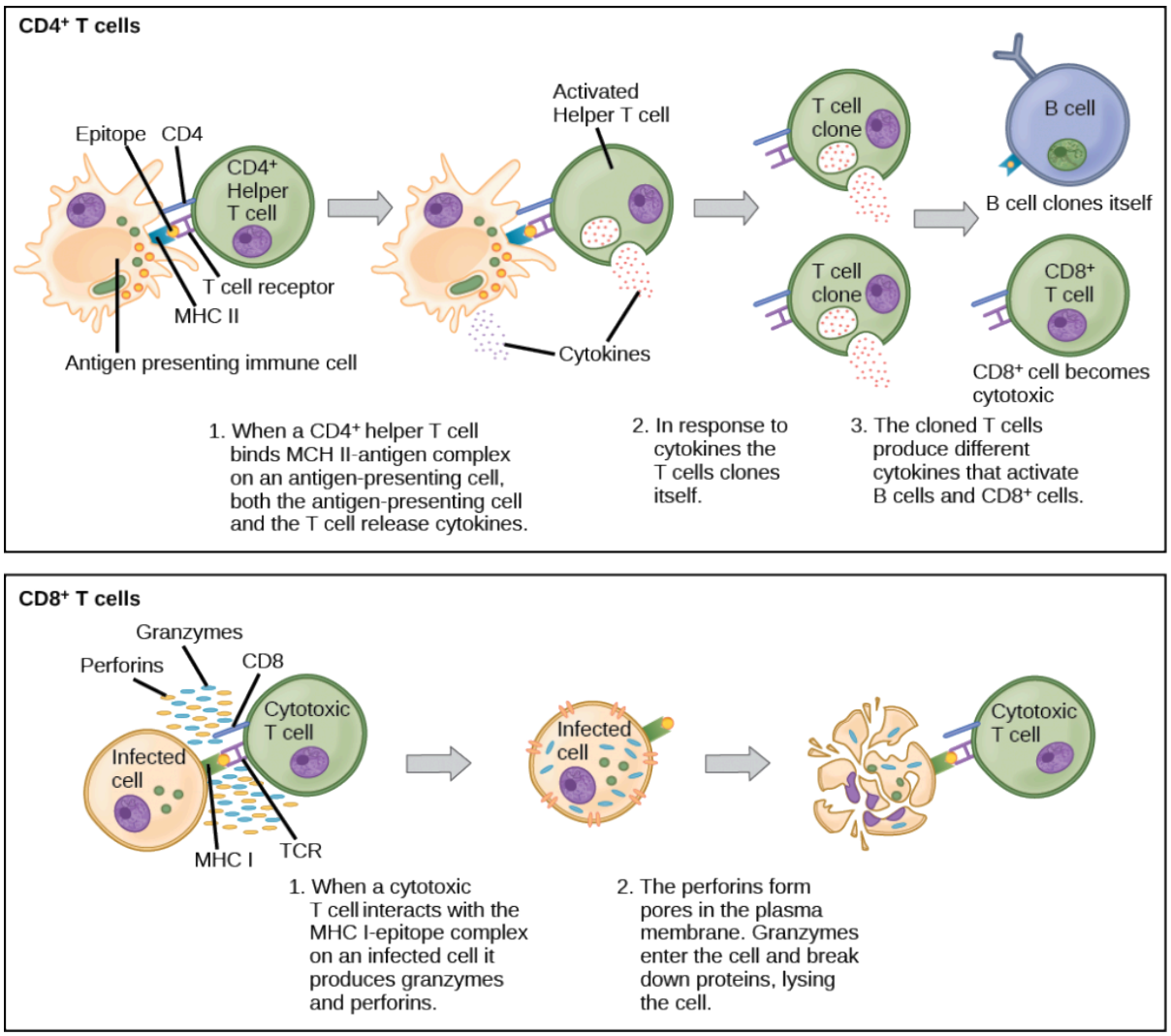


Figure 23.11. Naïve CD4+ T cells engage MHC II molecules on antigen-presenting cells (APCs) and become activated. Clones of the activated helper T cell, in turn, activate B cells and CD8+ T cells, which become cytotoxic T cells. Cytotoxic T cells kill infected cells.

Which of the following statements about T cells is false?

1. Helper T cells release cytokines while cytotoxic T cells kill the infected cell.
2. Helper T cells are $CD4^+$, while cytotoxic T cells are $CD8^+$.
3. MHC II is a receptor found on most body cells, while MHC I is a receptor found on immune cells only.
4. The T cell receptor is found on both $CD4^+$ and $CD8^+$ T cells.

Consider the innumerable possible antigens that an individual will be exposed to during a lifetime. The mammalian adaptive immune system is adept in responding appropriately to each antigen. Mammals have an enormous diversity of T cell populations, resulting from the diversity of TCRs. Each TCR consists of two polypeptide chains that span the T cell membrane, as illustrated in Figure 23.12; the chains are linked by a disulfide bridge. Each polypeptide chain is comprised of a constant domain and a variable domain: a domain, in this sense, is a specific region of a protein that may be regulatory or

structural. The intracellular domain is involved in intracellular signaling. A single T cell will express thousands of identical copies of one specific TCR variant on its cell surface. The specificity of the adaptive immune system occurs because it synthesizes millions of different T cell populations, each expressing a TCR that differs in its variable domain. This TCR diversity is achieved by the mutation and recombination of genes that encode these receptors in stem cell precursors of T cells. The binding between an antigen-displaying MHC molecule and a complementary TCR "match" indicates that the adaptive immune system needs to activate and produce that specific T cell because its structure is appropriate to recognize and destroy the invading pathogen.

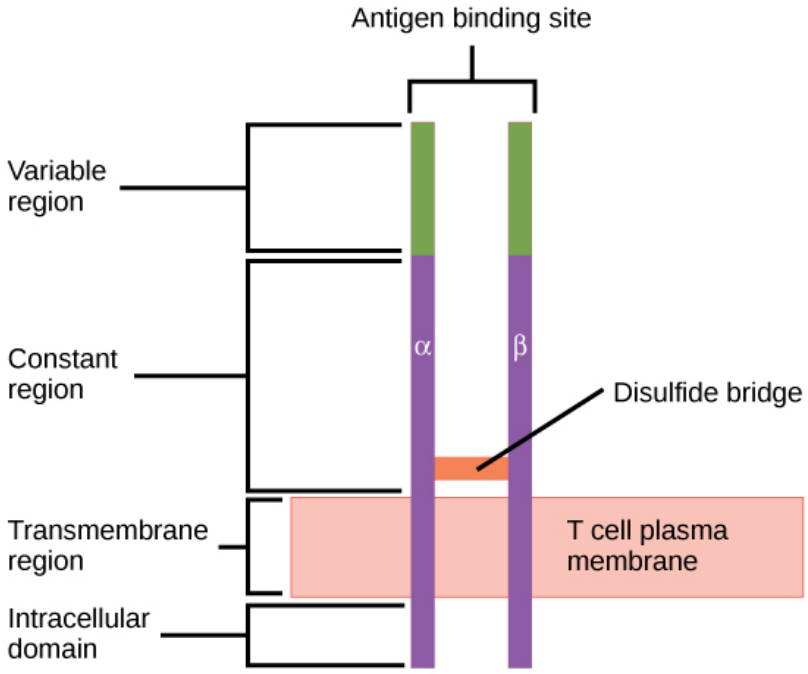


Figure 23.12. A T cell receptor spans the membrane and projects variable binding regions into the extracellular space to bind processed antigens via MHC molecules on APCs.

Helper T Lymphocytes

The T_H lymphocytes function indirectly to identify potential pathogens for other cells of the immune system. These cells are important for extracellular infections, such as those caused by certain bacteria, helminths, and protozoa. T_H lymphocytes recognize specific antigens displayed in the MHC II complexes of APCs. There are two major 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_H1 cells activate the action of cyotoxic T cells, as well as macrophages. T_H2 cells stimulate naïve B cells to destroy foreign invaders via antibody secretion. 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.

The T_H1-mediated response involves macrophages and is associated with inflammation. Recall the frontline defenses of macrophages involved in the innate immune response. Some intracellular bacteria, such as *Mycobacterium tuberculosis*, have evolved to multiply in macrophages after they have been engulfed. These pathogens evade attempts by macrophages to destroy and digest the pathogen. When

M. tuberculosis infection occurs, macrophages can stimulate naïve T cells to become T_H1 cells. These stimulated T cells secrete specific cytokines that send feedback to the macrophage to stimulate its digestive capabilities and allow it to destroy the colonizing *M. tuberculosis*. In the same manner, T_H1-activated macrophages also become better suited to ingest and kill tumor cells. In summary; T_H1 responses are directed toward intracellular invaders while T_H2 responses are aimed at those that are extracellular.

B Lymphocytes

When stimulated by the T_H2 pathway, naïve B cells differentiate into antibody-secreting plasma cells. A **plasma cell** is an immune cell that secrets antibodies; these cells arise from B cells that were stimulated by antigens. Similar to T cells, naïve B cells initially are coated in thousands of B cell receptors (BCRs), which are membrane-bound forms of Ig (immunoglobulin, or an antibody). The B cell receptor has two heavy chains and two light chains connected by disulfide linkages. Each chain has a constant and a variable region; the latter is involved in antigen binding. Two other membrane proteins, Ig alpha and Ig beta, are involved in signaling. The receptors of any particular B cell, as shown in Figure 23.13 are all the same, but the hundreds of millions of different B cells in an individual have distinct recognition domains that contribute to extensive diversity in the types of molecular structures to which they can bind. In this state, B cells function as APCs. They bind and engulf foreign antigens via their BCRs and then display processed antigens in the context of MHC II molecules to T_H2 cells. When a T_H2 cell detects that a B cell is bound to a relevant antigen, it secretes specific cytokines that induce the B cell to proliferate rapidly, which makes thousands of identical (clonal) copies of it, and then it synthesizes and secretes antibodies with the same antigen recognition pattern as the BCRs. The activation of B cells corresponding to one specific BCR variant and the dramatic proliferation of that variant is known as **clonal selection**. This phenomenon drastically, but briefly, changes the proportions of BCR variants expressed by the immune system, and shifts the balance toward BCRs specific to the infecting pathogen.

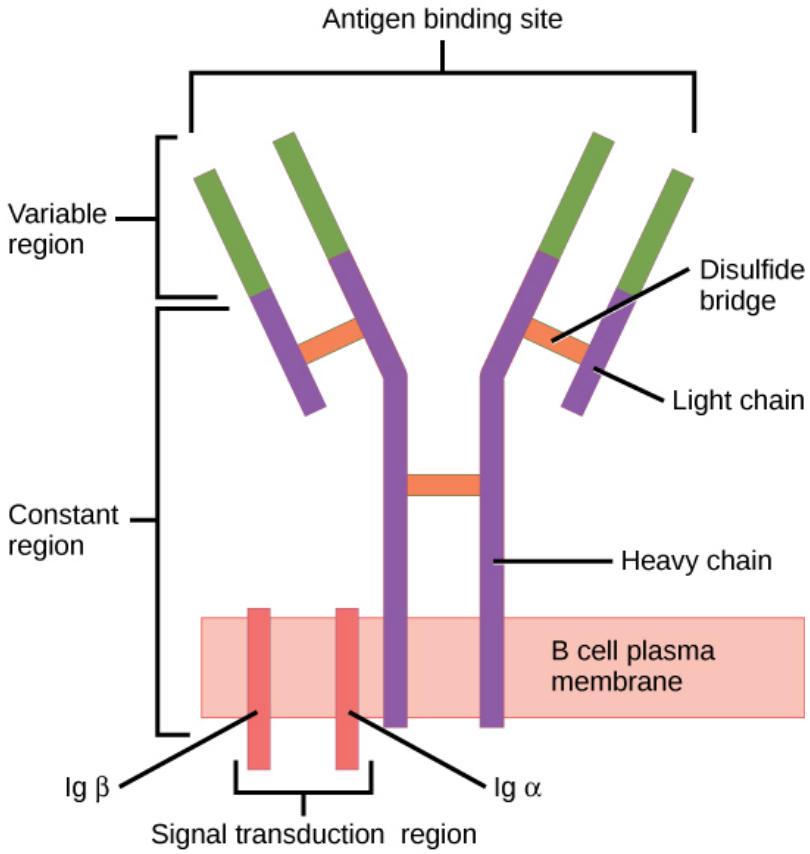


Figure 23.13. B cell receptors are embedded in the membranes of B cells and bind a variety of antigens through their variable regions. The signal transduction region transfers the signal into the cell.

T and B cells differ in one fundamental way: whereas T cells bind antigens that have been digested and embedded in MHC molecules by APCs, B cells function as APCs that bind intact antigens that have not been processed. Although T and B cells both react with molecules that are termed “antigens,” these lymphocytes actually respond to very different types of molecules. B cells must be able to bind intact antigens because they secrete antibodies that must recognize the pathogen directly, rather than digested remnants of the pathogen. Bacterial carbohydrate and lipid molecules can activate B cells independently from the T cells.

Cytotoxic T Lymphocytes

CTLs, a subclass of T cells, function to clear infections directly. The cell-mediated part of the adaptive immune system consists of CTLs that attack and destroy infected cells. CTLs 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. When APCs phagocytize pathogens and present MHC I-embedded antigens to naïve $CD8^+$ T cells that express complementary TCRs, the $CD8^+$ T cells become activated to proliferate according to clonal selection. These resulting CTLs then identify non-APCs displaying the same MHC I-embedded antigens (for example, viral proteins)—for example, the CTLs identify infected host cells.

Intracellularly, infected cells typically die after the infecting pathogen replicates to a sufficient

concentration and lyses the cell, as many viruses do. CTLs attempt to identify and destroy infected cells before the pathogen can replicate and escape, thereby halting the progression of intracellular infections. CTLs also support NK lymphocytes to destroy early cancers. Cytokines secreted by the T_H1 response that stimulates macrophages also stimulate CTLs and enhance their ability to identify and destroy infected cells and tumors.

CTLs sense MHC I-embedded antigens by directly interacting with infected cells via their TCRs. Binding of TCRs with antigens activates CTLs to release perforin and granzyme, degradative enzymes that will induce apoptosis of the infected cell. Recall that this is a similar destruction mechanism to that used by NK cells. In this process, the CTL does not become infected and is not harmed by the secretion of perforin and granzymes. In fact, the functions of NK cells and CTLs are complementary and maximize the removal of infected cells, as illustrated in Figure 23.14. If the NK cell cannot identify the "missing self" pattern of down-regulated MHC I molecules, then the CTL can identify it by the complex of MHC I with foreign antigens, which signals "altered self." Similarly, if the CTL cannot detect antigen-embedded MHC I because the receptors are depleted from the cell surface, NK cells will destroy the cell instead. CTLs also emit cytokines, such as interferons, that alter surface protein expression in other infected cells, such that the infected cells can be easily identified and destroyed. Moreover, these interferons can also prevent virally infected cells from releasing virus particles.

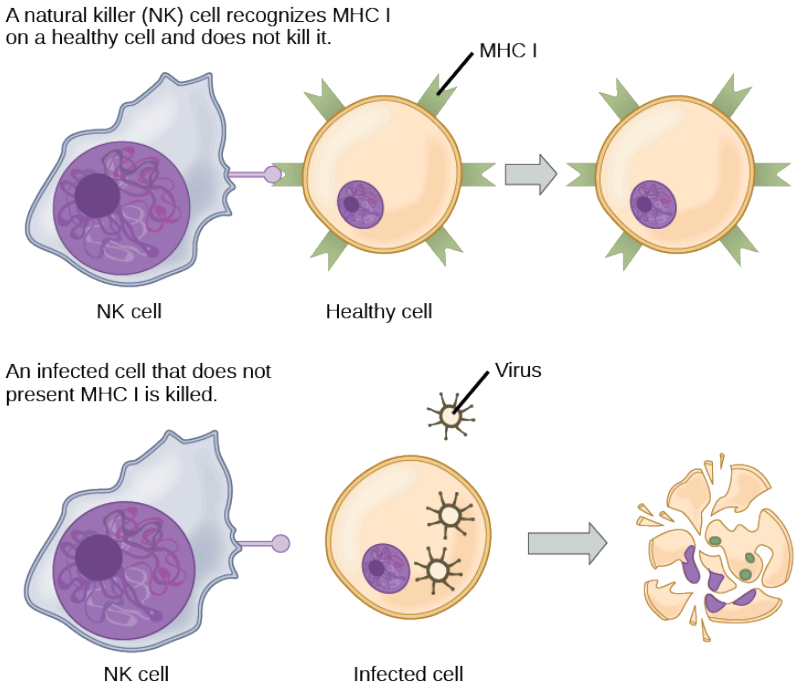


Figure 23.14. Natural killer (NK) cells recognize the MHC I receptor on healthy cells. If MHC I is absent, the cell is lysed.

Based on what you know about MHC receptors, why do you think an organ transplanted from an incompatible donor to a recipient will be rejected?

Plasma cells and CTLs are collectively called **effector cells**: they represent differentiated versions of their naïve counterparts, and they are involved in bringing about the immune defense of killing pathogens and infected host cells.

Mucosal Surfaces and Immune Tolerance

The innate and adaptive immune responses discussed thus far comprise the systemic immune system (affecting the whole body), which is distinct from the mucosal immune system. Mucosal immunity is formed by mucosa-associated lymphoid tissue, which functions independently of the systemic immune system, and which has its own innate and adaptive components. **Mucosa-associated lymphoid tissue (MALT)**, illustrated in Figure 23.15, is a collection of lymphatic tissue that combines with epithelial tissue lining the mucosa throughout the body. This tissue functions as the immune barrier and response in areas of the body with direct contact to the external environment. The systemic and mucosal immune systems use many of the same cell types. Foreign particles that make their way to MALT are taken up by absorptive epithelial cells called M cells and delivered to APCs located directly below the mucosal tissue. M cells function in the transport described, and are located in the Peyer's patch, a lymphoid nodule. APCs of the mucosal immune system are primarily dendritic cells, with B cells and macrophages having minor roles. Processed antigens displayed on APCs are detected by T cells in the MALT and at various mucosal induction sites, such as the tonsils, adenoids, appendix, or the mesenteric lymph nodes of the intestine. Activated T cells then migrate through the lymphatic system and into the circulatory system to mucosal sites of infection.

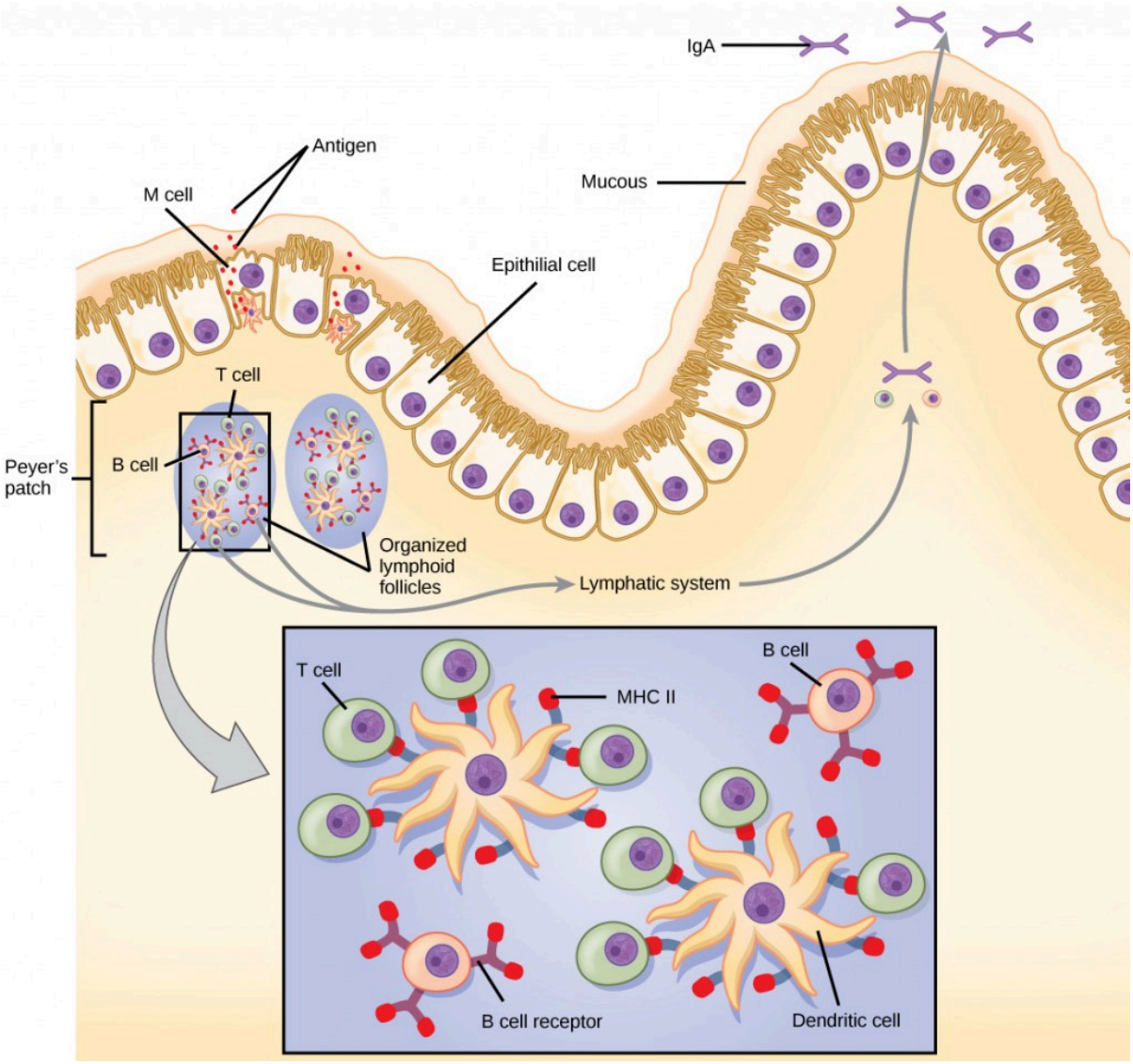


Figure 23.15. The topology and function of intestinal MALT is shown. Pathogens are taken up by M cells in the intestinal epithelium and excreted into a pocket formed by the inner surface of the cell. The pocket contains antigen-presenting cells such as dendritic cells, which engulf the antigens, then present them with MHC II molecules on the cell surface. The dendritic cells migrate to an underlying tissue called a Peyer's patch. Antigen-presenting cells, T cells, and B cells aggregate within the Peyer's patch, forming organized lymphoid follicles. There, some T cells and B cells are activated. Other antigen-loaded dendritic cells migrate through the lymphatic system where they activate B cells, T cells, and plasma cells in the lymph nodes. The activated cells then return to MALT tissue effector sites. IgA and other antibodies are secreted into the intestinal lumen.

MALT is a crucial component of a functional immune system because mucosal surfaces, such as the nasal passages, are the first tissues onto which inhaled or ingested pathogens are deposited. The mucosal tissue includes the mouth, pharynx, and esophagus, and the gastrointestinal, respiratory, and urogenital tracts.

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 is

described as **immune tolerance**. Immune tolerance is crucial for maintaining mucosal homeostasis given 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 an exceptionally 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. In addition to promoting immune tolerance of harmless antigens, other subsets of T_{reg} cells are involved in the prevention of the **autoimmune response**, which is an inappropriate immune response to host cells or self-antigens. Another T_{reg} class suppresses immune responses to harmful pathogens after the infection has cleared to minimize host cell damage induced by inflammation and cell lysis.

Immunological Memory

The adaptive immune system possesses a memory component that allows for an efficient and dramatic response upon reinvasion of the same pathogen. Memory is handled by the adaptive immune system with little reliance on cues from the innate response. During the adaptive immune response to a pathogen that has not been encountered before, called a primary 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, as illustrated in Figure 23.16.

A memory cell is an antigen-specific B or T lymphocyte that does not differentiate into effector cells during the primary immune response, but that can immediately become effector cells upon re-exposure to the same pathogen. During the primary immune response, memory cells do not respond to antigens and do not contribute to host defenses. 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.

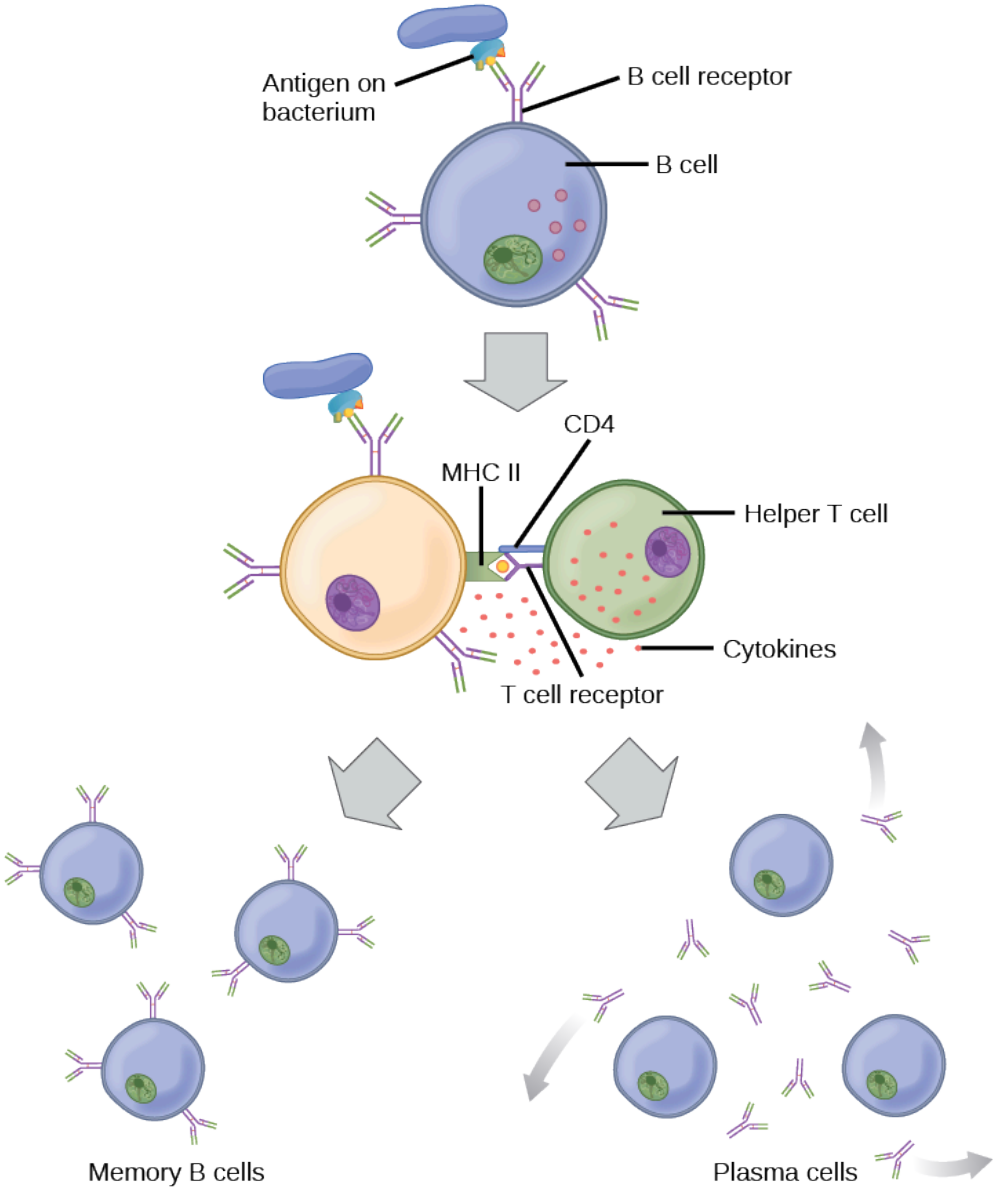


Figure 23.16. After initially binding an antigen to the B cell receptor (BCR), a B cell internalizes the antigen and presents it on MHC II. A helper T cell recognizes the MHC II–antigen complex and activates the B cell. As a result, memory B cells and plasma cells are made.

The Rh antigen is found on Rh-positive red blood cells. An Rh-negative female can usually carry an Rh-positive fetus to term without difficulty. However, if she has a second Rh-positive fetus, her body may launch an immune attack that causes hemolytic disease of the newborn. Why do you think hemolytic disease is only a problem during the second or subsequent pregnancies?

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 CTLs without input from APCs or T_H cells. One reason 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 and activated. Upon reinfection, this step is

skipped, and the result is a more rapid production of immune defenses. Memory B cells that differentiate into plasma cells output tens to hundreds-fold greater antibody amounts than were secreted during the primary response, as the graph in Figure 23.17 illustrates. 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.

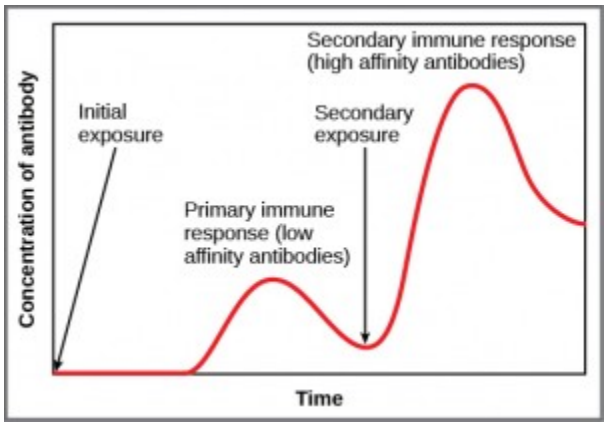


Figure 23.17. In the primary response to infection, antibodies are secreted first from plasma cells. Upon re-exposure to the same pathogen, memory cells differentiate into antibody-secreting plasma cells that output a greater amount of antibody for a longer period of time.

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, and because some memory cells die, certain vaccine courses involve one or more booster vaccinations to mimic repeat exposures: for instance, tetanus boosters are necessary every ten years because the memory cells only live that long.

Mucosal Immune Memory

A subset of T and B cells of the mucosal immune system differentiates into memory cells just as in the systemic immune system. Upon reinvasion of the same pathogen type, a pronounced immune response occurs at the mucosal site where the original pathogen deposited, but a collective defense is also organized within interconnected or adjacent mucosal tissue. For instance, the immune memory of an infection in the oral cavity would also elicit a response in the pharynx if the oral cavity was exposed to the same pathogen.

Vaccinologist

Vaccination (or immunization) involves the delivery, usually by injection as shown in Figure 23.18, of noninfectious antigen(s) derived from known pathogens. Other components, called adjuvants, are delivered in parallel to help stimulate the immune response. Immunological memory is the reason

vaccines work. Ideally, the effect of vaccination is to elicit immunological memory, and thus resistance to specific pathogens without the individual having to experience an infection.

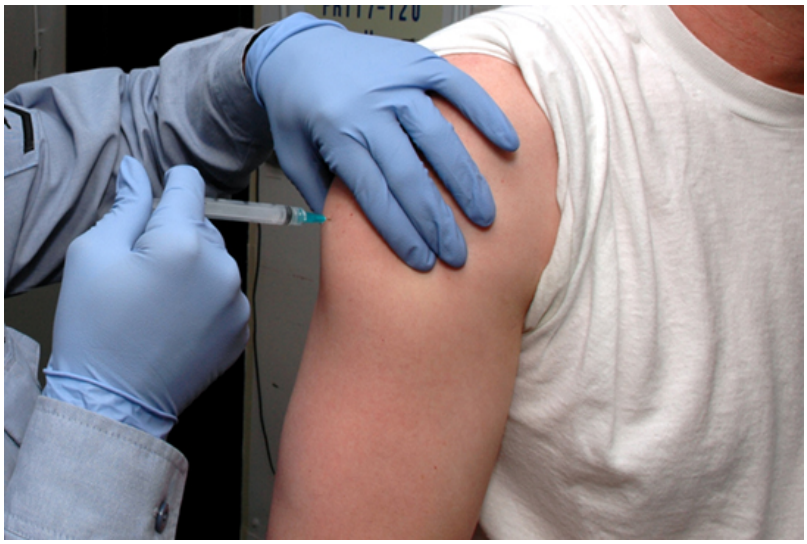


Figure 23.18. Vaccines are often delivered by injection into the arm. (credit: U.S. Navy Photographer's Mate Airman Apprentice Christopher D. Blachly)

Vaccinologists are involved in the process of vaccine development from the initial idea to the availability of the completed vaccine. This process can take decades, can cost millions of dollars, and can involve many obstacles along the way. For instance, injected vaccines stimulate the systemic immune system, eliciting humoral and cell-mediated immunity, but have little effect on the mucosal response, which presents a challenge because many pathogens are deposited and replicate in mucosal compartments, and the injection does not provide the most efficient immune memory for these disease agents. For this reason, vaccinologists are actively involved in developing new vaccines that are applied via intranasal, aerosol, oral, or transcutaneous (absorbed through the skin) delivery methods. Importantly, mucosal-administered vaccines elicit both mucosal and systemic immunity and produce the same level of disease resistance as injected vaccines.

Figure 23.19. The polio vaccine can be administered orally. (credit: modification of work by UNICEF Sverige)

Currently, a version of intranasal influenza vaccine is available, and the polio and typhoid vaccines can be administered orally, as shown in Figure 23.19. Similarly, the measles and rubella vaccines are being adapted to aerosol delivery using inhalation devices. Eventually, transgenic plants may be engineered to produce vaccine antigens that can be eaten to confer disease resistance. Other vaccines may be adapted to rectal or vaginal application to elicit immune responses in rectal, genitourinary, or reproductive mucosa. Finally, vaccine antigens may be adapted to transdermal application in which the skin is lightly scraped and microneedles are used to pierce the outermost layer. In addition to mobilizing the mucosal immune response, this new generation of vaccines may end the anxiety associated with injections and, in turn, improve patient participation.

Primary Centers of the Immune System

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 encompass monocytes (the precursor of macrophages) and lymphocytes. The majority of cells in the blood are erythrocytes (red blood cells). **Lymph** is a watery fluid that bathes tissues and organs with protective white blood cells and does not contain erythrocytes. 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).

The cells of the immune system originate from hematopoietic stem cells in the bone marrow. Cytokines stimulate these stem cells to differentiate into immune cells. B cell maturation occurs in the bone marrow, whereas naïve T cells transit from the bone marrow to the thymus for maturation. In the

thymus, immature T cells that express TCRs complementary to self-antigens are destroyed. This process helps prevent autoimmune responses.

On maturation, T and B lymphocytes circulate to various destinations. Lymph nodes scattered throughout the body, as illustrated in Figure 23.20, house large populations of T and B cells, dendritic cells, and macrophages. 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.

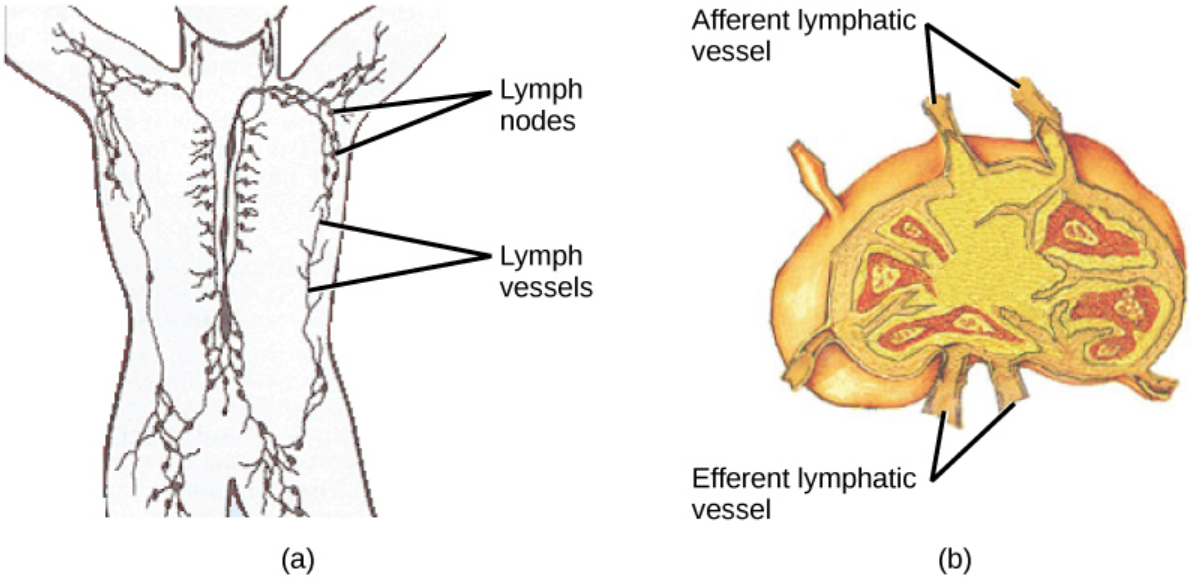


Figure 23.20. (a) Lymphatic vessels carry a clear fluid called lymph throughout the body. The liquid enters (b) lymph nodes through afferent vessels. Lymph nodes are filled with lymphocytes that purge infecting cells. The lymph then exits through efferent vessels. (credit: modification of work by NIH, NCI)

The spleen houses B and T cells, macrophages, dendritic cells, and NK cells. The spleen, shown in Figure 23.21, 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.

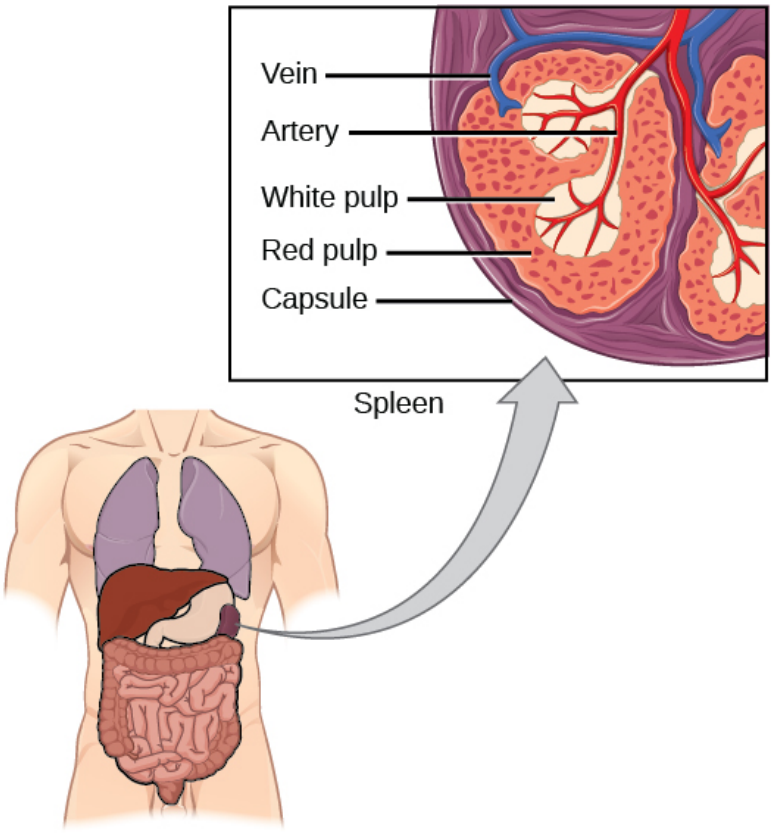


Figure 23.21. The spleen is similar to a lymph node but is much larger and filters blood instead of lymph. Blood enters the spleen through arteries and exits through veins. The spleen contains two types of tissue: red pulp and white pulp. Red pulp consists of cavities that store blood. Within the red pulp, damaged red blood cells are removed and replaced by new ones. White pulp is rich in lymphocytes that remove antigen-coated bacteria from the blood. (credit: modification of work by NCI)

Summary

The adaptive immune response is a slower-acting, longer-lasting, and more specific response than the innate response. However, the adaptive response requires information from the innate immune system to function. APCs display antigens via MHC molecules to complementary naïve T cells. In response, the T cells differentiate and proliferate, becoming T_H cells or CTLs. T_H cells stimulate B cells that have engulfed and presented pathogen-derived antigens. B cells differentiate into plasma cells that secrete antibodies, whereas CTLs induce apoptosis in intracellularly infected or cancerous cells. Memory cells persist after a primary exposure to a pathogen. If re-exposure occurs, memory cells differentiate into effector cells without input from the innate immune system. The mucosal immune system is largely independent from the systemic immune system but functions in a parallel fashion to protect the extensive mucosal surfaces of the body.

Exercises

1. Which of the following statements about T cells is false?
    1. Helper T cells release cytokines while cytotoxic T cells kill the infected cell.
    2. Helper T cells are CD4+, while cytotoxic T cells are $CD8^+$.
    3. MHC II is a receptor found on most body cells, while MHC I is a receptor found on immune cells only.
    4. The T cell receptor is found on both $CD4^+$ and $CD8^+$ T cells.
2. Based on what you know about MHC receptors, why do you think an organ transplanted from an incompatible donor to a recipient will be rejected?
3. The Rh antigen is found on Rh-positive red blood cells. An Rh-negative female can usually carry an Rh-positive fetus to term without difficulty. However, if she has a second Rh-positive fetus, her body may launch an immune attack that causes hemolytic disease of the newborn. Why do you think hemolytic disease is only a problem during the second or subsequent pregnancies?
4. Which of the following is both a phagocyte and an antigen-presenting cell?
    1. NK cell
    2. eosinophil
    3. neutrophil
    4. macrophage
5. Which immune cells bind MHC molecules on APCs via CD8 coreceptors on their cell surfaces?
    1. T_H cells
    2. CTLs
    3. mast cells
    4. basophils
6. What “self” pattern is identified by NK cells?
    1. altered self
    2. missing self
    3. normal self
    4. non-self
7. The acquired ability to prevent an unnecessary or destructive immune reaction to a harmless foreign particle, such as a food protein, is called _____.
    1. the T_H2 response
    2. allergy
    3. immune tolerance
    4. autoimmunity

8. A memory B cell can differentiate upon re-exposure to a pathogen of which cell type?
    1. CTL
    2. naïve B cell
    3. memory T cell
    4. plasma cell
9. Foreign particles circulating in the blood are filtered by the _____.
    1. spleen
    2. lymph nodes
    3. MALT
    4. lymph
10. Explain the difference between an epitope and an antigen.
11. What is a naïve B or T cell?
12. How does the T_H1 response differ from the T_H2 response?
13. In mammalian adaptive immune systems, T cell receptors are extraordinarily diverse. What function of the immune system results from this diversity, and how is this diversity achieved?
14. How do B and T cells differ with respect to antigens that they bind?
15. Why is the immune response after reinfection much faster than the adaptive immune response after the initial infection?

Answers

1. C
2. MHC receptors differ from person to person. Thus, MHC receptors on an incompatible donor are considered “non-self” and are rejected by the immune system.
3. If the blood of the mother and fetus mixes, memory cells that recognize the Rh antigen can form late in the first pregnancy. During subsequent pregnancies, these memory cells launch an immune attack on the fetal blood cells. Injection of anti-Rh antibody during the first pregnancy prevents the immune response from occurring.
4. D
5. B
6. B
7. C
8. D
9. A
10. An antigen is a molecule that reacts with some component of the immune response (antibody, B cell receptor, T cell receptor). An epitope is the region on the antigen through which binding with the immune component actually occurs.
11. A naïve T or B cell is one that has not been activated by binding to the appropriate epitope. Naïve

T and B cells cannot produce responses.

12. The T_H1 response involves the secretion of cytokines to stimulate macrophages and CTLs and improve their destruction of intracellular pathogens and tumor cells. It is associated with inflammation. The T_H2 response is involved in the stimulation of B cells into plasma cells that synthesize and secrete antibodies.
13. The diversity of TCRs allows the immune system to have millions of different T cells, and thereby to be specific in distinguishing antigens. This diversity arises from mutation and recombination in the genes that encode the variable regions of TCRs.
14. T cells bind antigens that have been digested and embedded in MHC molecules by APCs. In contrast, B cells function themselves as APCs to bind intact, unprocessed antigens.
15. Upon reinfection, the memory cells will immediately differentiate into plasma cells and CTLs without input from APCs or T_H cells. In contrast, the adaptive immune response to the initial infection requires time for naïve B and T cells with the appropriate antigen specificities to be identified and activated.

Glossary

adaptive immunity
immunity that has memory and occurs after exposure to an antigen either from a pathogen or a vaccination

antigen-presenting cell (APC)
immune cell that detects, engulfs, and informs the adaptive immune response about an infection by presenting the processed antigen on the cell surface

antigen
foreign or “non-self” protein that triggers the immune response

autoimmune response
inappropriate immune response to host cells or self-antigens

cell-mediated immune response
adaptive immune response that is carried out by T cells

clonal selection
activation of B cells corresponding to one specific BCR variant and the dramatic proliferation of that variant

cytotoxic T lymphocyte (CTL)
adaptive immune cell that directly kills infected cells via perforin and granzymes, and releases cytokines to enhance the immune response

dendritic cell
immune cell that processes antigen material and presents it on the surface of other cells to induce an immune response

effector cell
lymphocyte that has differentiated, such as a B cell, plasma cell, or cytotoxic T lymphocyte

epitope
small component of an antigen that is specifically recognized by antibodies, B cells, and T cells; the antigenic determinant

helper T lymphocyte (T_H)
cell of the adaptive immune system that binds APCs via MHC II molecules and stimulates B cells or secretes cytokines to initiate the immune response

humoral immune response
adaptive immune response that is controlled by activated B cells and antibodies

immune tolerance
acquired ability to prevent an unnecessary or harmful immune response to a detected foreign body known not to cause disease or to self-antigens

lymph
watery fluid that bathes tissues and organs with protective white blood cells and does not contain erythrocytes

memory cell
antigen-specific B or T lymphocyte that does not differentiate into effector cells during the primary immune response but that can immediately become an effector cell upon re-exposure to the same pathogen

mucosa-associated lymphoid tissue (MALT)
collection of lymphatic tissue that combines with epithelial tissue lining the mucosa throughout the body

regulatory T (T_{reg}) cell
specialized lymphocyte that suppresses local inflammation and inhibits the secretion of cytokines, antibodies, and other stimulatory immune factors; involved in immune tolerance

plasma cell is an immune cell that secrets antibodies; these cells arise from B cells that were stimulated by antigens.

23.3. Antibodies

Learning Objectives

By the end of this section, you will be able to:

- Explain cross-reactivity
- Describe the structure and function of antibodies
- Discuss antibody production

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 functional basis 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.

Antibody Structure

An antibody molecule is comprised of four polypeptides: two identical heavy chains (large peptide units) that are partially bound to each other in a “Y” formation, which are flanked by two identical light chains (small peptide units), as illustrated in

Figure 23.22. Bonds between the cysteine amino acids in the antibody molecule attach the polypeptides to each other. The areas where the antigen is recognized on the antibody are variable domains and the antibody base is composed of constant domains.

In germ-line B cells, the variable region of the light chain gene has 40 variable (V) and five joining (J) segments. An enzyme called DNA recombinase randomly excises most of these segments out of the gene, and splices one V segment to one J segment. During RNA processing, all but one V and J segment are spliced out. Recombination and splicing may result in over 10^6 possible VJ combinations. As a result, each differentiated B cell in the human body typically has a unique variable chain. The constant domain, which does not bind antibody, is the same for all antibodies.

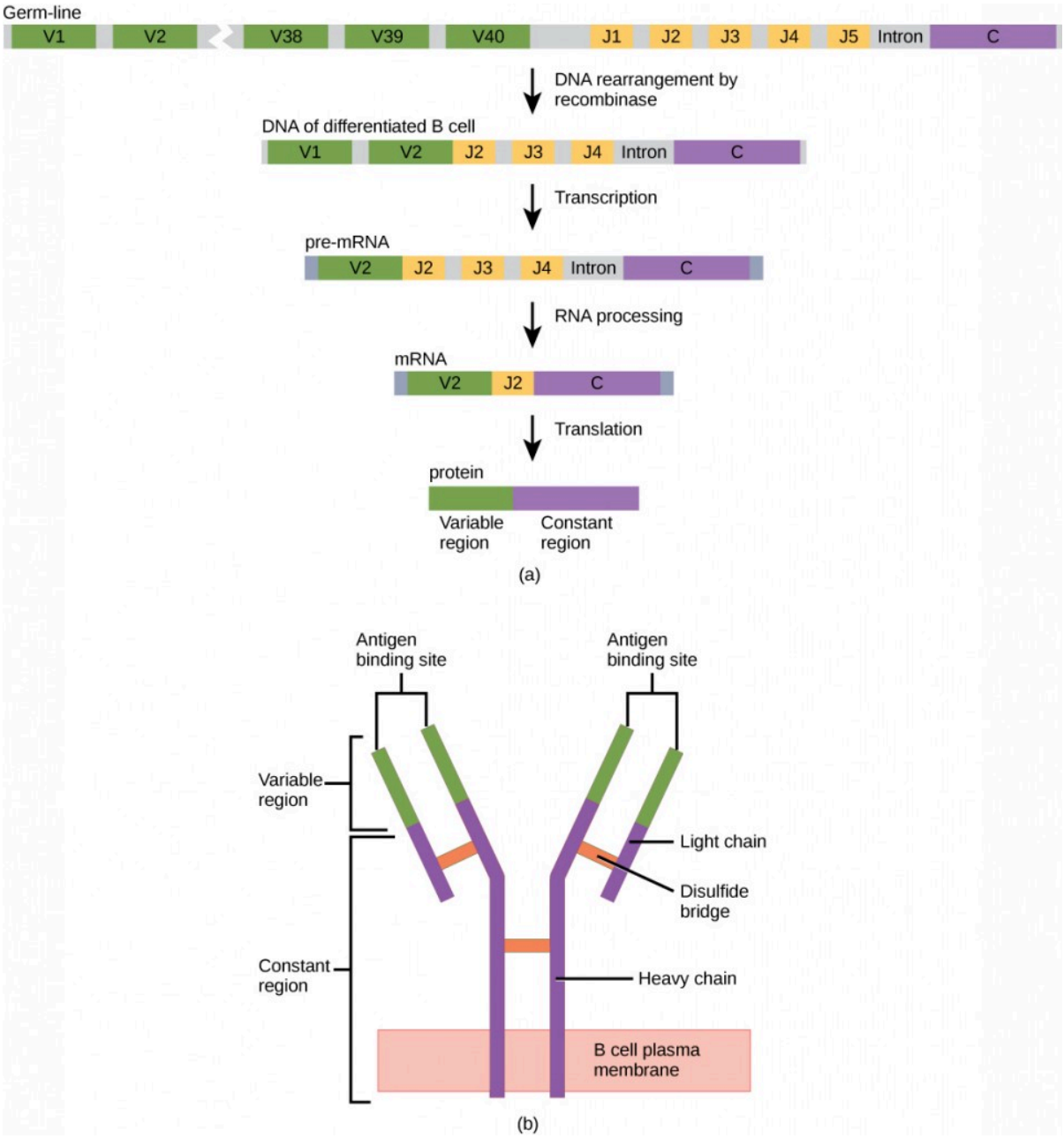


Figure 23.22. (a) As a germ-line B cell matures, an enzyme called DNA recombinase randomly excises V and J segments from the light chain gene. Splicing at the mRNA level results in further gene rearrangement. As a result, (b) each antibody has a unique variable region capable of binding a different antigen.

Similar to TCRs and BCRs, antibody diversity is produced by the mutation and recombination of approximately 300 different gene segments encoding the light and heavy chain variable domains in precursor cells that are destined to become B cells. The variable domains from the heavy and light chains interact to form the binding site through which an antibody can bind a specific epitope on an antigen. The numbers of repeated constant domains in Ig classes are the same for all antibodies corresponding to a specific class. Antibodies are structurally similar to the extracellular component of the BCRs, and B cell maturation to plasma cells can be visualized in simple terms as the cell acquires the ability to secrete the extracellular portion of its BCR in large quantities.

Antibody Classes

Antibodies can be divided into five classes—IgM, IgG, IgA, IgD, IgE—based on their physiochemical, structural, and immunological properties. IgGs, which make up about 80 percent of all antibodies, have heavy chains that consist of one variable domain and three identical constant domains. IgA and IgD also have three constant domains per heavy chain, whereas IgM and IgE each have four constant domains per heavy chain. The variable domain determines binding specificity and the constant domain of the heavy chain determines the immunological mechanism of action of the corresponding antibody class. It is possible for two antibodies to have the same binding specificities but be in different classes and, therefore, to be involved in different functions.

After an adaptive defense is produced against a pathogen, typically plasma cells first secrete IgM into the blood. BCRs on naïve B cells are of the IgM class and occasionally IgD class. IgM molecules make up approximately ten percent of all antibodies. Prior to antibody secretion, plasma cells assemble IgM molecules into pentamers (five individual antibodies) linked by a joining (J) chain, as shown in Figure 23.23. The pentamer arrangement means that these macromolecules can bind ten identical antigens. However, IgM molecules released early in the adaptive immune response do not bind to antigens as stably as IgGs, which are one of the possible types of antibodies secreted in large quantities upon re-exposure to the same pathogen. Figure 23.23 summarizes the properties of immunoglobulins and illustrates their basic structures.

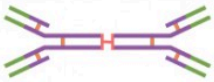
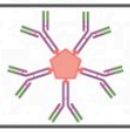
Name	Properties	Structure
IgA	Found in mucous, saliva, tears, and breast milk. Protects against pathogens.	
IgD	Part of the B cell receptor. Activates basophils and mast cells.	
IgE	Protects against parasitic worms. Responsible for allergic reactions.	
IgG	Secreted by plasma cells in the blood. Able to cross the placenta into the fetus.	
IgM	May be attached to the surface of a B cell or secreted into the blood. Responsible for early stages of immunity.	

Figure 23.23. Immunoglobulins have different functions, but all are composed of light and heavy chains that form a Y-shaped structure.

IgAs populate the saliva, tears, breast milk, and mucus secretions of the gastrointestinal, respiratory, and genitourinary tracts. Collectively, these bodily fluids coat and protect the extensive mucosa (4000 square feet in humans). The total number of IgA molecules in these bodily secretions is greater than the number of IgG molecules in the blood serum. A small amount of IgA is also secreted into the serum in monomeric form. Conversely, some IgM is secreted into bodily fluids of the mucosa. Similar to IgM,

IgA molecules are secreted as polymeric structures linked with a J chain. However, IgAs are secreted mostly as dimeric molecules, not pentamers.

IgE is present in the serum in small quantities and is best characterized in its role as an allergy mediator. IgD is also present in small quantities. Similar to IgM, BCRs of the IgD class are found on the surface of naïve B cells. This class supports antigen recognition and maturation of B cells to plasma cells.

Antibody Functions

Differentiated plasma cells are crucial players in the humoral response, and the antibodies they secrete are particularly significant against extracellular pathogens and toxins. Antibodies circulate freely and act independently of plasma cells. Antibodies can be transferred from one individual to another to temporarily protect against infectious disease. For instance, a person who has recently produced a successful immune response against a particular disease agent can donate blood to a nonimmune recipient and confer temporary immunity through antibodies in the donor's blood serum. This phenomenon is called **passive immunity**; it also occurs naturally during breastfeeding, which makes breastfed infants highly resistant to infections during the first few months of life.

Antibodies coat extracellular pathogens and neutralize them, as illustrated in Figure 23.24, by blocking key sites on the pathogen that enhance their infectivity (such as receptors that "dock" pathogens on host cells). Antibody neutralization can prevent pathogens from entering and infecting host cells, as opposed to the CTL-mediated approach of killing cells that are already infected to prevent progression of an established infection. The neutralized antibody-coated pathogens can then be filtered by the spleen and eliminated in urine or feces.

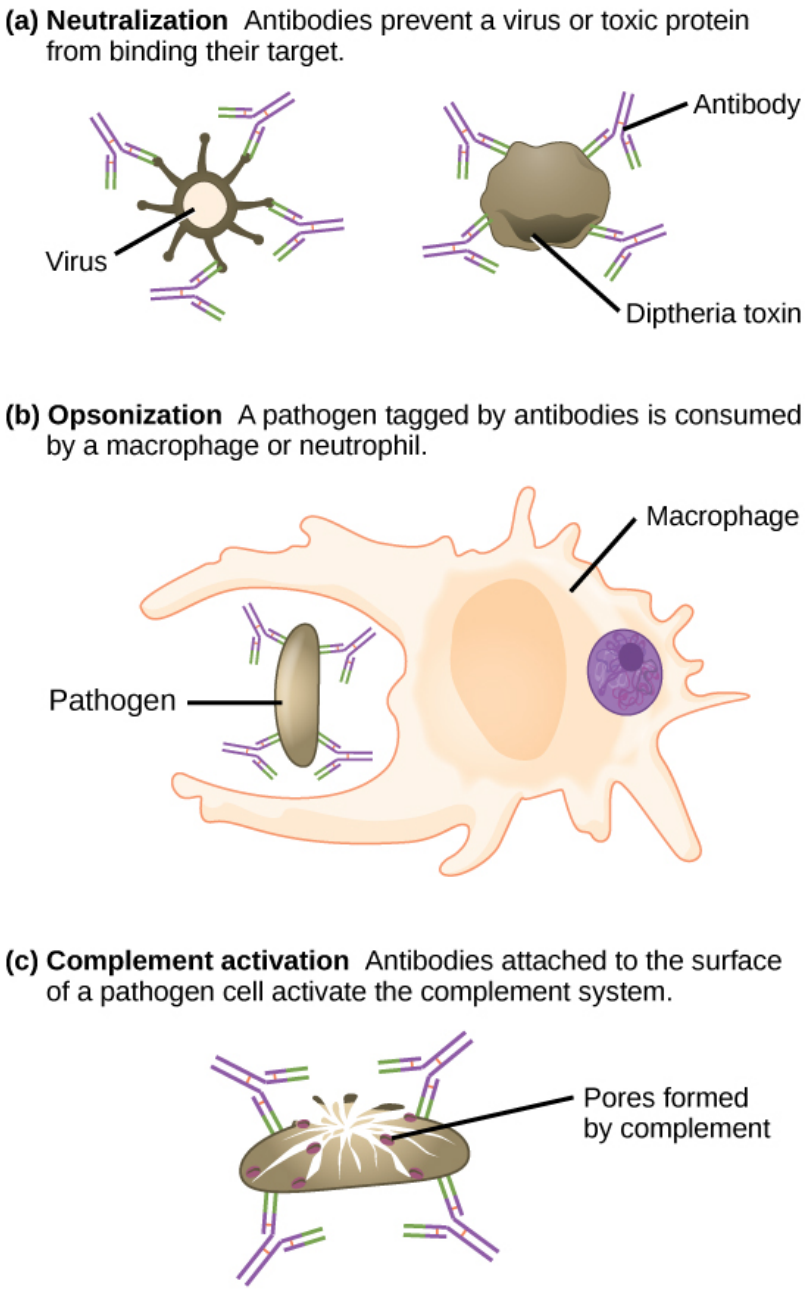


Figure 23.24. Antibodies may inhibit infection by (a) preventing the antigen from binding its target, (b) tagging a pathogen for destruction by macrophages or neutrophils, or (c) activating the complement cascade.

Antibodies also mark pathogens for destruction by phagocytic cells, such as macrophages or neutrophils, because phagocytic cells are highly attracted to macromolecules complexed with antibodies. Phagocytic enhancement by antibodies is called opsonization. In a process called complement fixation, IgM and IgG in serum bind to antigens and provide docking sites onto which sequential complement proteins can bind. The combination of antibodies and complement enhances opsonization even further and promotes rapid clearing of pathogens.

Affinity, Avidity, and Cross Reactivity

Not all antibodies bind with the same strength, specificity, and stability. In fact, antibodies exhibit different **affinities** (attraction) depending on the molecular complementarity between antigen and antibody molecules, as illustrated in Figure 23.25. An antibody with a higher affinity for a particular antigen would bind more strongly and stably, and thus would be expected to present a more challenging defense against the pathogen corresponding to the specific antigen.

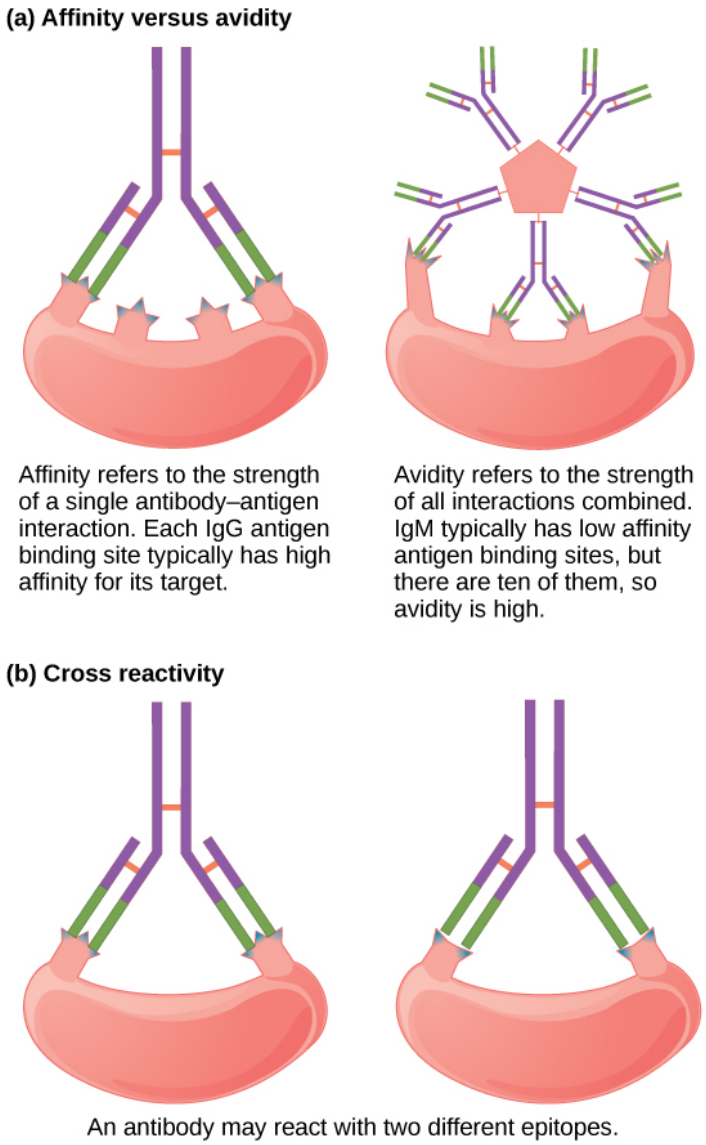


Figure 23.25. (a) Affinity refers to the strength of single interaction between antigen and antibody, while avidity refers to the strength of all interactions combined. (b) An antibody may cross react with different epitopes.

The term **avidity** describes binding by antibody classes that are secreted as joined, multivalent structures (such as IgM and IgA). Although avidity measures the strength of binding, just as affinity does, the avidity is not simply the sum of the affinities of the antibodies in a multimeric structure. The avidity depends on the number of identical binding sites on the antigen being detected, as well as other

physical and chemical factors. Typically, multimeric antibodies, such as pentameric IgM, are classified as having lower affinity than monomeric antibodies, but high avidity. Essentially, the fact that multimeric antibodies can bind many antigens simultaneously balances their slightly lower binding strength for each antibody/antigen interaction.

Antibodies secreted after binding to one epitope on an antigen may exhibit cross reactivity for the same or similar epitopes on different antigens. Because an epitope corresponds to such a small region (the surface area of about four to six amino acids), it is possible for different macromolecules to exhibit the same molecular identities and orientations over short regions. **Cross reactivity** describes when an antibody binds not to the antigen that elicited its synthesis and secretion, but to a different antigen.

Cross reactivity can be beneficial if an individual develops immunity to several related pathogens despite having only been exposed to or vaccinated against one of them. For instance, antibody cross reactivity may occur against the similar surface structures of various Gram-negative bacteria. Conversely, antibodies raised against pathogenic molecular components that resemble self molecules may incorrectly mark host cells for destruction and cause autoimmune damage. Patients who develop systemic lupus erythematosus (SLE) commonly exhibit antibodies that react with their own DNA. These antibodies may have been initially raised against the nucleic acid of microorganisms but later cross-reacted with self-antigens. This phenomenon is also called molecular mimicry.

Antibodies of the Mucosal Immune System

Antibodies synthesized by the mucosal immune system include IgA and IgM. Activated B cells differentiate into mucosal plasma cells that synthesize and secrete dimeric IgA, and to a lesser extent, pentameric IgM. Secreted IgA is abundant in tears, saliva, breast milk, and in secretions of the gastrointestinal and respiratory tracts. Antibody secretion results in a local humoral response at epithelial surfaces and prevents infection of the mucosa by binding and neutralizing pathogens.

Summary

Antibodies (immunoglobulins) are the molecules secreted from plasma cells that mediate the humoral immune response. There are five antibody classes; an antibody's class determines its mechanism of action and production site but does not control its binding specificity. Antibodies bind antigens via variable domains and can either neutralize pathogens or mark them for phagocytosis or activate the complement cascade.

Exercises

1. The structure of an antibody is similar to the extracellular component of which receptor?
    1. MHC I
    2. MHC II
    3. BCR
    4. none of the above

2. The first antibody class to appear in the serum in response to a newly encountered pathogen is _____.
    1. IgM
    2. IgA
    3. IgG
    4. IgE
3. What is the most abundant antibody class detected in the serum upon reexposure to a pathogen or in reaction to a vaccine?
    1. IgM
    2. IgA
    3. IgG
    4. IgE
4. Breastfed infants typically are resistant to disease because of _____.
    1. active immunity
    2. passive immunity
    3. immune tolerance
    4. immune memory
5. What are the benefits and costs of antibody cross reactivity?

Answers

1. C
2. A
3. C
4. B
5. Cross reactivity of antibodies can be beneficial when it allows an individual's immune system to respond to an array of similar pathogens after being exposed to just one of them. A potential cost of cross reactivity is an antibody response to parts of the body (self) in addition to the appropriate antigen.

Glossary

affinity
attraction of molecular complementarity between antigen and antibody molecules

antibody
protein that is produced by plasma cells after stimulation by an antigen; also known as an

immunoglobulin

avidity
total binding strength of a multivalent antibody with antigen

cross reactivity
binding of an antibody to an epitope corresponding to an antigen that is different from the one the antibody was raised against

passive immunity
transfer of antibodies from one individual to another to provide temporary protection against pathogensplasma cellimmune cell that secrets antibodies; these cells arise from B cells that were stimulated by antigens

23.4. Disruptions in the Immune System

Learning Objectives

By the end of this section, you will be able to:

- Describe hypersensitivity
- Define autoimmunity

A functioning immune system is essential for survival, but even the sophisticated cellular and molecular defenses of the mammalian immune response can be defeated by pathogens at virtually every step. In the competition between immune protection and pathogen evasion, pathogens have the advantage of more rapid evolution because of their shorter generation time and other characteristics. For instance, *Streptococcus pneumoniae* (bacterium that cause pneumonia and meningitis) surrounds itself with a capsule that inhibits phagocytes from engulfing it and displaying antigens to the adaptive immune system. *Staphylococcus aureus* (bacterium that can cause skin infections, abscesses, and meningitis) synthesizes a toxin called leukocidin that kills phagocytes after they engulf the bacterium. Other pathogens can also hinder the adaptive immune system. HIV infects T_H cells via their CD4 surface molecules, gradually depleting the number of T_H cells in the body; this inhibits the adaptive immune system's capacity to generate sufficient responses to infection or tumors. As a result, HIV-infected individuals often suffer from infections that would not cause illness in people with healthy immune systems but which can cause devastating illness to immune-compromised individuals. Maladaptive responses of immune cells and molecules themselves can also disrupt the proper functioning of the entire system, leading to host cell damage that could become fatal.

Immunodeficiency

Failures, insufficiencies, or delays at any level of the immune response can allow pathogens or tumor cells to gain a foothold and replicate or proliferate to high enough levels that the immune system becomes overwhelmed. **Immunodeficiency** is the failure, insufficiency, or delay in the response of the immune system, which may be acquired or inherited. Immunodeficiency can be acquired as a result of infection with certain pathogens (such as HIV), chemical exposure (including certain medical treatments), malnutrition, or possibly by extreme stress. For instance, radiation exposure can destroy populations of lymphocytes and elevate an individual's susceptibility to infections and cancer. Dozens of genetic disorders result in immunodeficiencies, including Severe Combined Immunodeficiency (SCID), Bare lymphocyte syndrome, and MHC II deficiencies. Rarely, primary immunodeficiencies that are present from birth may occur. Neutropenia is one form in which the immune system produces a below-average number of neutrophils, the body's most abundant phagocytes. As a result, bacterial infections may go unrestricted in the blood, causing serious complications.

Hypersensitivities

Maladaptive immune responses toward harmless foreign substances or self antigens that occur after tissue sensitization are termed **hypersensitivities**. The types of hypersensitivities include immediate, delayed, and autoimmunity. A large proportion of the population is affected by one or more types of hypersensitivity.

Allergies

The immune reaction that results from immediate hypersensitivities in which an antibody-mediated immune response occurs within minutes of exposure to a harmless antigen is called an **allergy**

. In the United States, 20 percent of the population exhibits symptoms of allergy or asthma, whereas 55 percent test positive against one or more allergens. Upon initial exposure to a potential allergen, an allergic individual synthesizes antibodies of the IgE class via the typical process of APCs presenting processed antigen to T_H cells that stimulate B cells to produce IgE. This class of antibodies also mediates the immune response to parasitic worms. The constant domain of the IgE molecules interact with mast cells embedded in connective tissues. This process primes, or sensitizes, the tissue. Upon subsequent exposure to the same allergen, IgE molecules on mast cells bind the antigen via their variable domains and stimulate the mast cell to release the modified amino acids histamine and serotonin; these chemical mediators then recruit eosinophils which mediate allergic responses. Figure 23.26 shows an example of an allergic response to ragweed pollen. The effects of an allergic reaction range from mild symptoms like sneezing and itchy, watery eyes to more severe or even life-threatening reactions involving intensely itchy welts or hives, airway contraction with severe respiratory distress, and plummeting blood pressure. This extreme reaction is known as anaphylactic shock. If not treated with epinephrine to counter the blood pressure and breathing effects, this condition can be fatal.

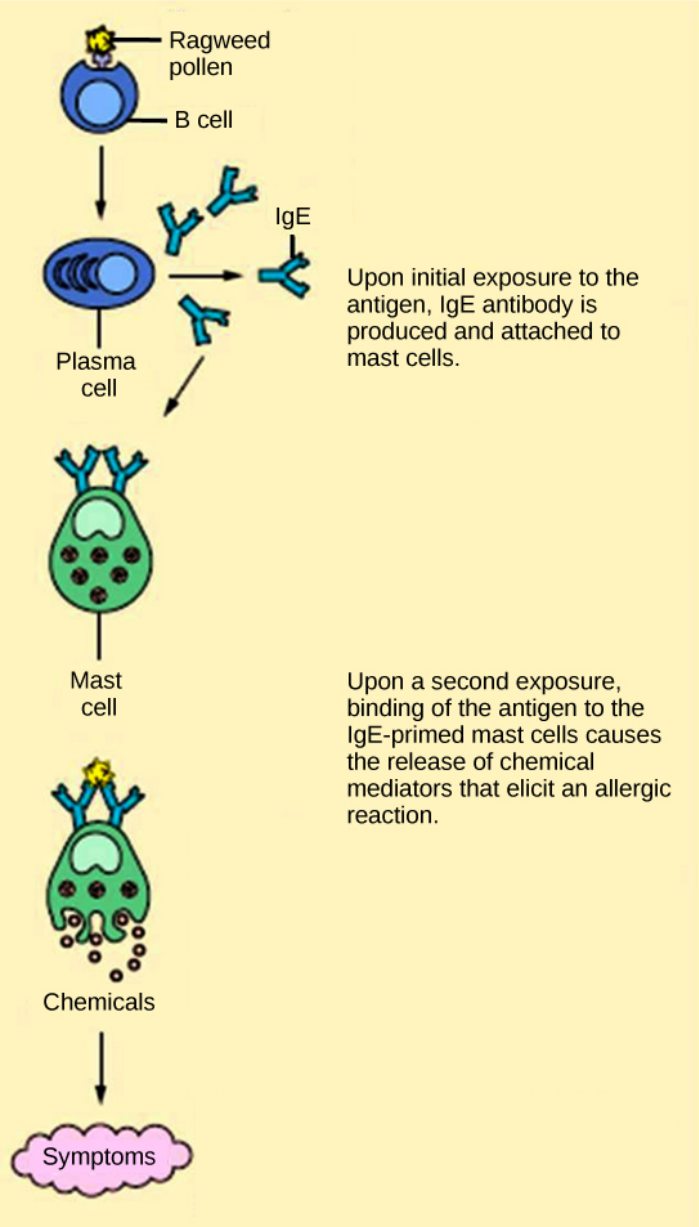


Figure 23.26. On first exposure to an allergen, an IgE antibody is synthesized by plasma cells in response to a harmless antigen. The IgE molecules bind to mast cells, and on secondary exposure, the mast cells release histamines and other modulators that affect the symptoms of allergy. (credit: modification of work by NIH)

Delayed hypersensitivity is a cell-mediated immune response that takes approximately one to two days after secondary exposure for a maximal reaction to be observed. This type of hypersensitivity involves the T_H1 cytokine-mediated inflammatory response and may manifest as local tissue lesions or contact dermatitis (rash or skin irritation). Delayed hypersensitivity occurs in some individuals in response to contact with certain types of jewelry or cosmetics. Delayed hypersensitivity facilitates the immune response to poison ivy and is also the reason why the skin test for tuberculosis results in a small region of inflammation on individuals who were previously exposed to *Mycobacterium tuberculosis*. That is also why cortisone is used to treat such responses: it will inhibit cytokine production.

Autoimmunity

Autoimmunity is a type of hypersensitivity to self antigens that affects approximately five percent of the population. Most types of autoimmunity involve the humoral immune response. Antibodies that inappropriately mark self components as foreign are termed **autoantibodies**. In patients with the autoimmune disease myasthenia gravis, muscle cell receptors that induce contraction in response to acetylcholine are targeted by antibodies. The result is muscle weakness that may include marked difficultly with fine and/or gross motor functions. In systemic lupus erythematosus, a diffuse autoantibody response to the individual's own DNA and proteins results in various systemic diseases. As illustrated in Figure 23.27, systemic lupus erythematosus may affect the heart, joints, lungs, skin, kidneys, central nervous system, or other tissues, causing tissue damage via antibody binding, complement recruitment, lysis, and inflammation.

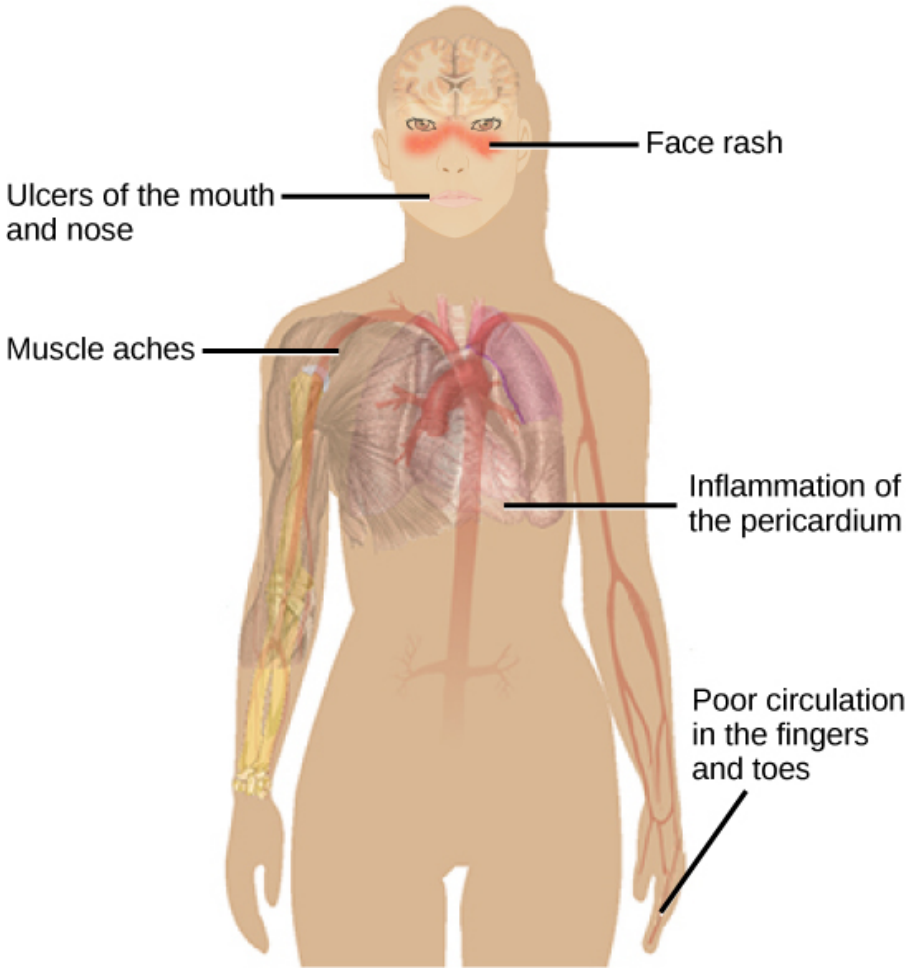


Figure 23.27. Systemic lupus erythematosus is characterized by autoimmunity to the individual's own DNA and/or proteins, which leads to varied dysfunction of the organs. (credit: modification of work by Mikael Häggström)

Autoimmunity can develop with time, and its causes may be rooted in molecular mimicry. Antibodies and TCRs may bind self antigens that are structurally similar to pathogen antigens, which the immune receptors first raised. As an example, infection with *Streptococcus pyogenes* (bacterium that causes strep throat) may generate antibodies or T cells that react with heart muscle, which has a similar

structure to the surface of *S. pyogenes*. These antibodies can damage heart muscle with autoimmune attacks, leading to rheumatic fever. Insulin-dependent (Type 1) diabetes mellitus arises from a destructive inflammatory T_H1 response against insulin-producing cells of the pancreas. Patients with this autoimmunity must be injected with insulin that originates from other sources.

Summary

Immune disruptions may involve insufficient immune responses or inappropriate immune targets. Immunodeficiency increases an individual's susceptibility to infections and cancers. Hypersensitivities are misdirected responses either to harmless foreign particles, as in the case of allergies, or to host factors, as in the case of autoimmunity. Reactions to self components may be the result of molecular mimicry.

Exercises

1. Allergy to pollen is classified as:
    1. an autoimmune reaction
    2. immunodeficiency
    3. delayed hypersensitivity
    4. immediate hypersensitivity
2. A potential cause of acquired autoimmunity is _____.
    1. tissue hypersensitivity
    2. molecular mimicry
    3. histamine release
    4. radiation exposure
3. Autoantibodies are probably involved in:
    1. reactions to poison ivy
    2. pollen allergies
    3. systemic lupus erythematosus
    4. HIV/AIDS
4. Which of the following diseases is not due to autoimmunity?
    1. rheumatic fever
    2. systemic lupus erythematosus
    3. diabetes mellitus
    4. HIV/AIDS

Answers

1. D
2. B
3. C
4. D

Glossary

allergy
immune reaction that results from immediate hypersensitivities in which an antibody-mediated immune response occurs within minutes of exposure to a harmless antigen

autoantibody
antibody that incorrectly marks “self” components as foreign and stimulates the immune response

autoimmunity
type of hypersensitivity to self antigens

hypersensitivities
spectrum of maladaptive immune responses toward harmless foreign particles or self antigens; occurs after tissue sensitization and includes immediate-type (allergy), delayed-type, and autoimmunity

immunodeficiency
failure, insufficiency, or delay at any level of the immune system, which may be acquired or inherited

Chapter 24. Animal Reproduction and Development

Figure 24.1. Female seahorses produce eggs for reproduction that are then fertilized by the male. Unlike almost all other animals, the male seahorse then gestates the young until birth. (credit: modification of work by "cliff1066"/Flickr)

Introduction

Animal reproduction is necessary for the survival of a species. In the animal kingdom, there are innumerable ways that species reproduce. Asexual reproduction produces genetically identical organisms (clones), whereas in sexual reproduction, the genetic material of two individuals combines to produce offspring that are genetically different from their parents. During sexual reproduction the male gamete (sperm) may be placed inside the female's body for internal fertilization, or the sperm and eggs may be released into the environment for external fertilization. Seahorses, like the one shown in Figure 24.1, provide an example of the latter. Following a mating dance, the female lays eggs in the male seahorse's abdominal brood pouch where they are fertilized. The eggs hatch and the offspring develop in the pouch for several weeks.

24.1. Reproduction Methods

Learning Objectives

By the end of this section, you will be able to:

- Describe advantages and disadvantages of asexual and sexual reproduction
- Discuss asexual reproduction methods
- Discuss sexual reproduction methods

Animals produce offspring through asexual and/or 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. 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 asexually-reproducing species may be at a disadvantage because all the offspring are genetically identical and may not have the genetic variation to survive in new or different conditions. On the other hand, the rapid rates of asexual reproduction may allow for a speedy response to environmental changes if individuals have mutations. An additional advantage of asexual reproduction is that colonization of new habitats may be easier when an individual does not need to find a mate to reproduce.

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 species a better chance of surviving in an unpredictable or changing environment. Species that reproduce sexually must maintain two different types of individuals, males and females, which can limit the ability to colonize new habitats as both sexes must be present.

Asexual Reproduction

Asexual reproduction occurs in prokaryotic microorganisms (bacteria) and in some eukaryotic single-celled and multi-celled organisms. There are a number of ways that animals reproduce asexually.

Fission

Fission, also called binary fission, occurs in prokaryotic microorganisms and in some invertebrate, multi-celled organisms. After a period of growth, an organism splits into two separate organisms. Some unicellular eukaryotic organisms undergo binary fission by mitosis. In other organisms, part of the individual separates and forms a second individual. This process occurs, for example, in many asteroid echinoderms through splitting of the central disk. Some sea anemones and some coral polyps (Figure 24.2) also reproduce through fission.

Figure 24.2. Coral polyps reproduce asexually by fission. (credit: G. P. Schmahl, NOAA FGBNMS Manager)

Budding

Budding is a form of asexual reproduction that results from the outgrowth of a part of a cell or body region leading to a separation from the original organism into two individuals. Budding occurs commonly in some invertebrate animals such as corals and hydras. In hydras, a bud forms that develops into an adult and breaks away from the main body, as illustrated in Figure 24.3, whereas in coral budding, the bud does not detach and multiplies as part of a new colony.

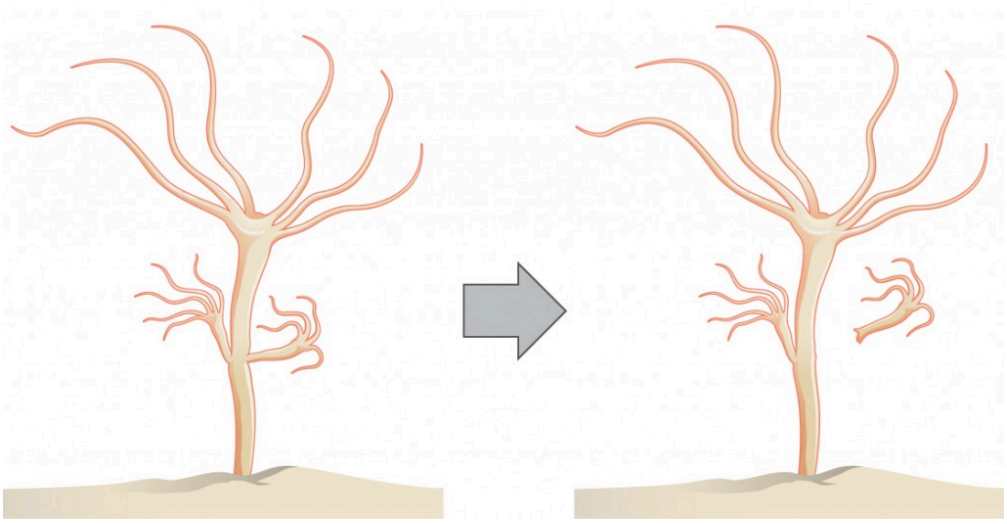


Figure 24.3. Hydra reproduce asexually through budding.

Concept in Action

Watch a video of a hydra budding.

Fragmentation

Fragmentation is the breaking of the body into two parts with subsequent regeneration. If the animal is capable of fragmentation, and the part is big enough, a separate individual will regrow.

For example, in many sea stars, asexual reproduction is accomplished by fragmentation. Figure 24.4 illustrates a sea star for which an arm of the individual is broken off and regenerates a new sea star. 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. Fragmentation also occurs in annelid worms, turbellarians, and poriferans.

Figure 24.4. Sea stars can reproduce through fragmentation. The large arm, a fragment from another sea star, is developing into a new individual.

Note that in fragmentation, there is generally a noticeable difference in the size of the individuals, whereas in fission, two individuals of approximate size are formed.

Parthenogenesis

Parthenogenesis is a form of asexual reproduction where an egg develops into a complete individual without being fertilized. The resulting offspring can be either haploid or diploid, depending on the process and the species. Parthenogenesis occurs in invertebrates such as water flees, rotifers, aphids, stick insects, some ants, wasps, and bees. Bees use parthenogenesis to produce haploid males (drones) and diploid females (workers). If an egg is fertilized, a queen is produced. The queen bee controls the reproduction of the hive bees to regulate the type of bee produced.

Some vertebrate animals—such as certain reptiles, amphibians, and fish—also reproduce through parthenogenesis. Although more common in plants, parthenogenesis has been observed in animal species that were segregated by sex in terrestrial or marine zoos. Two female Komodo dragons, a hammerhead shark, and a blacktop shark have produced parthenogenic young when the females have been isolated from males.

Sexual Reproduction

Sexual reproduction is the combination of (usually haploid) reproductive cells from two individuals to form a third (usually diploid) unique offspring. Sexual reproduction produces offspring with novel combinations of genes. This can be an adaptive advantage in unstable or unpredictable environments. As humans, we are used to thinking of animals as having two separate sexes—male and female—determined at conception. However, in the animal kingdom, there are many variations on this theme.

Hermaphroditism

Hermaphroditism occurs in animals where one individual has both male and female reproductive parts. Invertebrates such as earthworms, slugs, tapeworms and snails, shown in Figure 24.5, are often hermaphroditic. Hermaphrodites may self-fertilize or may mate with another of their species, fertilizing each other and both producing offspring. Self fertilization is common in animals that have limited mobility or are not motile, such as barnacles and clams.

Figure 24.5. Many snails are hermaphrodites. When two individuals mate, they can produce up to one hundred eggs each. (credit: Assaf Shtilman)

Sex Determination

Mammalian sex determination is determined genetically by the presence of X and Y chromosomes. Individuals homozygous for X (XX) are female and heterozygous individuals (XY) are male. The presence of a Y chromosome causes the development of male characteristics and its absence results in female characteristics. The XY system is also found in some insects and plants.

Avian sex determination is dependent on the presence of Z and W chromosomes. Homozygous for Z (ZZ) results in a male and heterozygous (ZW) results in a female. The W appears to be essential in determining the sex of the individual, similar to the Y chromosome in mammals. Some fish, crustaceans, insects (such as butterflies and moths), and reptiles use this system.

The sex of some species is not determined by genetics but by some aspect of the environment. Sex determination in some crocodiles and turtles, for example, is often dependent on the temperature during critical periods of egg development. This is referred to as environmental sex determination, or more specifically as temperature-dependent sex determination. In many turtles, cooler temperatures during egg incubation produce males and warm temperatures produce females. In some crocodiles, moderate temperatures produce males and both warm and cool temperatures produce females. In some species, sex is both genetic- and temperature-dependent.

Individuals of some species change their sex during their lives, alternating between male and female. If the individual is female first, it is termed protogyny or “first female,” if it is male first, its termed protandry or “first male.” Oysters, for example, are born male, grow, and become female and lay eggs; some oyster species change sex multiple times.

Summary

Reproduction may be asexual when one individual produces genetically identical offspring, or sexual when the genetic material from two individuals is combined to produce genetically diverse offspring. Asexual reproduction occurs through fission, budding, and fragmentation. Sexual reproduction may mean the joining of sperm and eggs within animals' bodies or it may mean the release of sperm and eggs into the environment. An individual may be one sex, or both; it may start out as one sex and switch during its life, or it may stay male or female.

Exercises

1. Which form of reproduction is thought to be best in a stable environment?
    1. asexual
    2. sexual
    3. budding
    4. parthenogenesis
2. Which form of reproduction can result from damage to the original animal?
    1. asexual
    2. fragmentation
    3. budding
    4. parthenogenesis
3. Which form of reproduction is useful to an animal with little mobility that reproduces sexually?
    1. fission
    2. budding
    3. parthenogenesis
    4. hermaphroditism
4. Genetically unique individuals are produced through _____.
    1. sexual reproduction
    2. parthenogenesis
    3. budding
    4. fragmentation
5. Why is sexual reproduction useful if only half the animals can produce offspring and two separate cells must be combined to form a third?
6. What determines which sex will result in offspring of birds and mammals?

Answers

1. A
2. B
3. D
4. A
5. Sexual reproduction produces a new combination of genes in the offspring that may better enable them to survive changes in the environment and assist in the survival of the species.
6. The presence of the W chromosome in birds determines femaleness and the presence of the Y chromosome in mammals determines maleness. The absence of those chromosomes and the homogeneity of the offspring (ZZ or XX) leads to the development of the other sex.

Glossary

asexual reproduction
form of reproduction that produces offspring that are genetically identical to the parent

budding
form of asexual reproduction that results from the outgrowth of a part of a cell leading to a separation from the original animal into two individuals

fission
(also, binary fission) method by which multicellular organisms increase in size or asexual reproduction in which a unicellular organism splits into two separate organisms by mitosis

fragmentation
cutting or fragmenting of the original animal into parts and the growth of a separate animal from each part

hermaphroditism
state of having both male and female reproductive parts within the same individual

parthenogenesis
form of asexual reproduction where an egg develops into a complete individual without being fertilized

sexual reproduction
mixing of genetic material from two individuals to produce genetically unique offspring

24.2. Fertilization

Learning Objectives

By the end of this section, you will be able to:

- Discuss internal and external methods of fertilization
- Describe the methods used by animals for development of offspring during gestation
- Describe the anatomical adaptions that occurred in animals to facilitate reproduction

Sexual reproduction starts with the combination of a sperm and an egg in 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 seahorse 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 release of the reproductive material may be triggered by 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).

Figure 24.6 shows salmon spawning in a shallow stream. Frogs, like those shown in Figure 24.7, corals, mayflies, and mosquitoes also spawn.

Figure 24.6. Salmon reproduce through spawning. (credit: Dan Bennett)

Figure 24.7. During sexual reproduction in toads, the male grasps the female from behind and externally fertilizes the eggs as they are deposited. (credit: "OakleyOriginals"/Flickr)

Pairs of fish that are not broadcast spawners may exhibit courtship behavior. This allows the female to select a particular male. The trigger for egg and sperm release (spawning) causes the egg and sperm to be placed in a small area, enhancing the possibility of fertilization.

External fertilization in an aquatic environment protects the eggs from drying out. Broadcast spawning can result in a greater mixture of the genes within a group, leading to higher genetic diversity and a greater chance of species survival in a hostile environment. For sessile aquatic organisms like sponges, broadcast spawning is the only mechanism for fertilization and colonization of new environments. The presence of the fertilized eggs and developing young in the water provides opportunities for predation resulting in a loss of offspring. Therefore, millions of eggs must be produced by individuals, and the offspring produced through this method must mature rapidly. The survival rate of eggs produced through broadcast spawning is low.

Internal Fertilization

Internal fertilization occurs most often in land-based animals, although some aquatic animals also use this method. 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. This occurs in most bony fish, many reptiles, some cartilaginous fish, most amphibians, two mammals, and all birds. Reptiles and insects produce leathery eggs, while birds and turtles produce eggs with high concentrations of calcium carbonate in the shell, making them hard. Chicken eggs are an example of this second type.

In **ovoviparity**, fertilized eggs are retained in the female, but the embryo obtains its nourishment from the egg's yolk and the young are fully developed when they are hatched. This occurs in some bony fish (like the guppy *Lebistes reticulatus*), some sharks, some lizards, some snakes (such as the garter snake *Thamnophis sirtalis*), some vipers, and some invertebrate animals (like the Madagascar hissing cockroach *Gromphadorhina portentosa*).

In **viviparity** the young develop within the female, receiving nourishment from the mother's blood through a placenta. The offspring develops in the female and is born alive. This occurs in most mammals, some cartilaginous fish, and a few reptiles.

Internal fertilization has the advantage of protecting the fertilized egg from dehydration on land. The embryo is isolated within the female, which limits predation on the young. Internal fertilization enhances the fertilization of eggs by a specific male. Fewer offspring are produced through this method, but their survival rate is higher than that for external fertilization.

The Evolution of Reproduction

Once multicellular organisms evolved and developed specialized cells, some also developed tissues and organs with specialized functions. An early development in reproduction occurred in the Annelids. These organisms produce sperm and eggs from undifferentiated cells in their coelom and store them in that cavity. When the coelom becomes filled, the cells are released through an excretory opening or by the body splitting open. Reproductive organs evolved with the development of gonads that produce sperm and eggs. These cells went through meiosis, an adaption of mitosis, which reduced the number of chromosomes in each reproductive cell by half, while increasing the number of cells through cell division.

Complete reproductive systems were developed in insects, with separate sexes. Sperm are made in

testes and then travel through coiled tubes to the epididymis for storage. Eggs mature in the ovary. When they are released from the ovary, they travel to the uterine tubes for fertilization. Some insects have a specialized sac, called a **spermatheca**, which stores sperm for later use, sometimes up to a year. Fertilization can be timed with environmental or food conditions that are optimal for offspring survival.

Vertebrates have similar structures, with a few differences. Non-mammals, such as birds and reptiles, have a common body opening, called a **cloaca**, for the digestive, excretory and reproductive systems. Coupling between birds usually involves positioning the cloaca openings opposite each other for transfer of sperm. Mammals have separate openings for the systems in the female and a uterus for support of developing offspring. The uterus has two chambers in species that produce large numbers of offspring at a time, while species that produce one offspring, such as primates, have a single uterus.

Sperm transfer from the male to the female during reproduction ranges from releasing the sperm into the watery environment for external fertilization, to the joining of cloaca in birds, to the development of a penis for direct delivery into the female's vagina in mammals.

Summary

Sexual reproduction starts with the combination of a sperm and an egg in a process called fertilization. This can occur either outside the bodies or inside the female. Both methods have advantages and disadvantages. Once fertilized, the eggs can develop inside the female or outside. If the egg develops outside the body, it usually has a protective covering over it. Animal anatomy evolved various ways to fertilize, hold, or expel the egg. The method of fertilization varies among animals. Some species release the egg and sperm into the environment, some species retain the egg and receive the sperm into the female body and then expel the developing embryo covered with shell, while still other species retain the developing offspring through the gestation period.

Exercises

1. External fertilization occurs in which type of environment?
    1. aquatic
    2. forested
    3. savanna
    4. steppe
2. Which term applies to egg development within the female with nourishment derived from a yolk?
    1. oviparity
    2. viviparity
    3. ovoviparity
    4. ovovoparity
3. Which term applies to egg development outside the female with nourishment derived from a yolk?

1. oviparity
2. viviparity
3. ovoviparity
4. ovovoparity

4. What are the advantages and disadvantages of external and internal forms of fertilization?
5. Why would paired external fertilization be preferable to group spawning?

Answers

1. A
2. C
3. A
4. External fertilization can create large numbers of offspring without requiring specialized delivery or reproductive support organs. Offspring develop and mature quickly compared to internally fertilizing species. A disadvantage is that the offspring are out in the environment and predation can account for large loss of offspring. The embryos are susceptible to changes in the environment, which further depletes their numbers. Internally fertilizing species control their environment and protect their offspring from predators but must have specialized organs to complete these tasks and usually produce fewer embryos.
5. Paired external fertilization allows the female to select the male for mating. It also has a greater chance of fertilization taking place, whereas spawning just puts a large number of sperm and eggs together and random interactions result in the fertilization.

Glossary

cloaca
common body opening for the digestive, excretory, and reproductive systems found in non-mammals, such as birds

external fertilization
fertilization of egg by sperm outside animal body, often during spawning

internal fertilization
fertilization of egg by sperm inside the body of the female

oviparity
process by which fertilized eggs are laid outside the female's body and develop there, receiving nourishment from the yolk that is a part of the egg

ovoviparity
process by which fertilized eggs are retained within the female; the embryo obtains its nourishment from the egg's yolk and the young are fully developed when they are hatched

spermatheca
specialized sac that stores sperm for later use

viviparity
process in which the young develop within the female, receiving nourishment from the mother’s blood through a placenta

24.3. Human Reproductive Anatomy and Gametogenesis

Learning Objectives

By the end of this section, you will be able to:

- Describe human male and female reproductive anatomies
- Discuss the human sexual response
- Describe spermatogenesis and oogenesis and discuss their differences and similarities

As animals became more complex, specific organs and organ systems developed to support specific functions for the organism. The reproductive structures that evolved in land animals allow males and females to mate, fertilize internally, and support the growth and development of offspring.

Human Reproductive Anatomy

The reproductive tissues of male and female humans develop similarly *in utero* until a low level of the hormone testosterone is released from male gonads. Testosterone causes the undeveloped tissues to differentiate into male sexual organs. When testosterone is absent, the tissues develop into female sexual tissues. Primitive gonads become testes or ovaries. Tissues that produce a penis in males produce a clitoris in females. The tissue that will become the scrotum in a male becomes the labia in a female; that is, they are homologous structures.

Male Reproductive Anatomy

In the male reproductive system, the **scrotum** houses the testicles or testes (singular: testis), including providing passage for blood vessels, nerves, and muscles related to testicular function. The **testes** are a pair of male reproductive organs that produce sperm and some reproductive hormones. Each testis is approximately 2.5 by 3.8 cm (1.5 by 1 in) in size and divided into wedge-shaped lobules by connective tissue called septa. Coiled in each wedge are seminiferous tubules that produce sperm.

Sperm are immobile at body temperature; therefore, the scrotum and penis are external to the body, as illustrated in Figure 24.8 so that a proper temperature is maintained for motility. In land mammals, the pair of testes must be suspended outside the body at about 2° C lower than body temperature to produce viable sperm. Infertility can occur in land mammals when the testes do not descend through the abdominal cavity during fetal development.

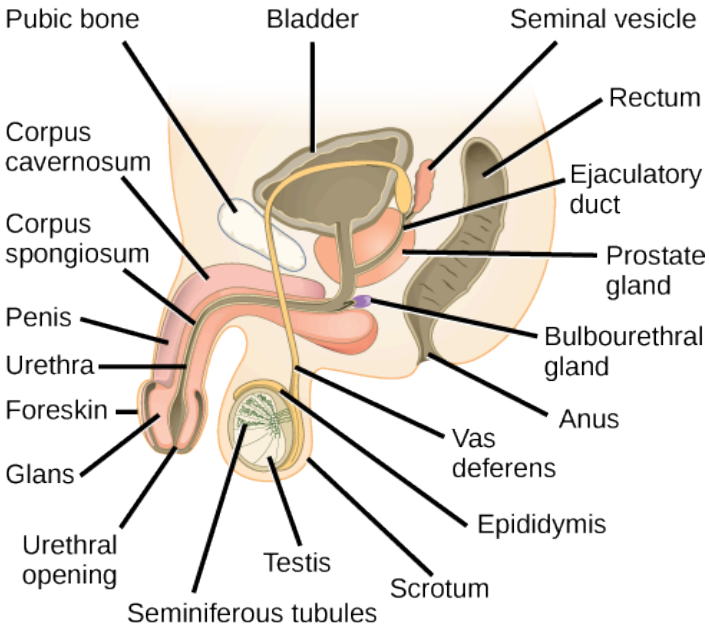


Figure 24.8. The reproductive structures of the human male are shown.

Which of the following statements about the male reproductive system is false?

1. The vas deferens carries sperm from the testes to the penis.
2. Sperm mature in seminiferous tubules in the testes.
3. Both the prostate and the bulbourethral glands produce components of the semen.
4. The prostate gland is located in the testes.

Sperm mature in **seminiferous tubules** that are coiled inside the testes, as illustrated in Figure 24.8. The walls of the seminiferous tubules are made up of the developing sperm cells, with the least developed sperm at the periphery of the tubule and the fully developed sperm in the lumen. The sperm cells are mixed with “nursemaid” cells called Sertoli cells which protect the germ cells and promote their development. Other cells mixed in the wall of the tubules are the interstitial cells of Leydig. These cells produce high levels of testosterone once the male reaches adolescence.

When the sperm have developed flagella and are nearly mature, they leave the testicles and enter the epididymis, shown in Figure 24.8. This structure resembles a comma and lies along the top and posterior portion of the testes; it is the site of sperm maturation. The sperm leave the epididymis and enter the vas deferens (or ductus deferens), which carries the sperm, behind the bladder, and forms the ejaculatory duct with the duct from the seminal vesicles. During a vasectomy, a section of the vas deferens is removed, preventing sperm from being passed out of the body during ejaculation and preventing fertilization.

Semen is a mixture of sperm and spermatic duct secretions (about 10 percent of the total) and fluids from accessory glands that contribute most of the semen’s volume. Sperm are haploid cells, consisting of a flagellum as a tail, a neck that contains the cell’s energy-producing mitochondria, and a head that contains the genetic material. Figure 24.9 shows a micrograph of human sperm as well as a diagram of the parts of the sperm. An acrosome is found at the top of the head of the sperm. This structure contains

lysosomal enzymes that can digest the protective coverings that surround the egg to help the sperm penetrate and fertilize the egg. An ejaculate will contain from two to five milliliters of fluid with from 50–120 million sperm per milliliter.

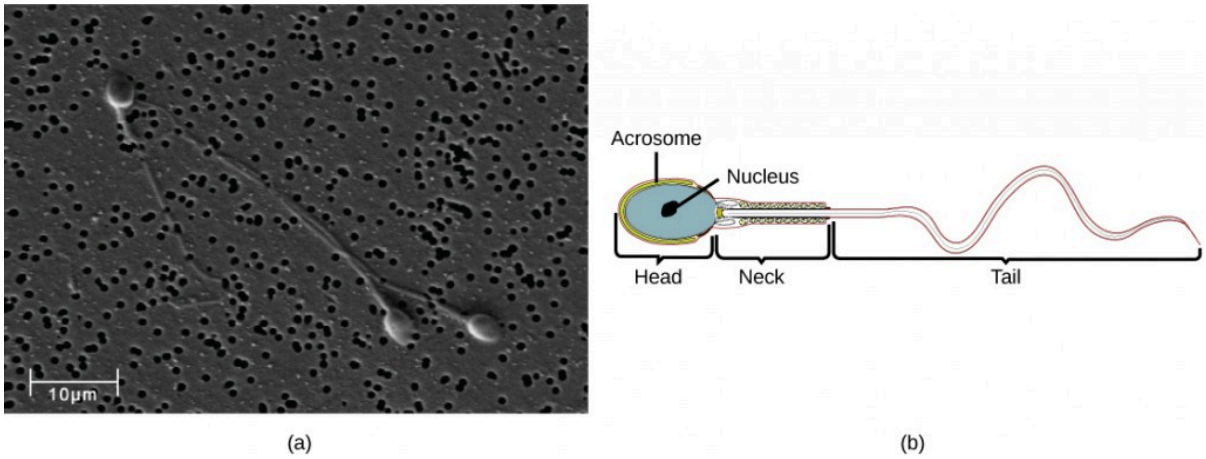


Figure 24.9. Human sperm, visualized using scanning electron microscopy, have a flagellum, neck, and head. (credit b: modification of work by Mariana Ruiz Villareal; scale-bar data from Matt Russell)

The bulk of the semen comes from the accessory glands associated with the male reproductive system. These are the seminal vesicles, the prostate gland, and the bulbourethral gland, all of which are illustrated in Figure 24.8. The **seminal vesicles** are a pair of glands that lie along the posterior border of the urinary bladder. The glands make a solution that is thick, yellowish, and alkaline. As sperm are only motile in an alkaline environment, a basic pH is important to reverse the acidity of the vaginal environment. The solution also contains mucus, fructose (a sperm mitochondrial nutrient), a coagulating enzyme, ascorbic acid, and local-acting hormones called prostaglandins. The seminal vesicle glands account for 60 percent of the bulk of semen.

The **penis**, illustrated in Figure 24.8, is an organ that drains urine from the renal bladder and functions as a copulatory organ during intercourse. The penis contains three tubes of erectile tissue running through the length of the organ. These consist of a pair of tubes on the dorsal side, called the corpus cavernosum, and a single tube of tissue on the ventral side, called the corpus spongiosum. This tissue will become engorged with blood, becoming erect and hard, in preparation for intercourse. The organ is inserted into the vagina culminating with an ejaculation. During intercourse, the smooth muscle sphincters at the opening to the renal bladder close and prevent urine from entering the penis. An orgasm is a two-stage process: first, glands and accessory organs connected to the testes contract, then semen (containing sperm) is expelled through the urethra during ejaculation. After intercourse, the blood drains from the erectile tissue and the penis becomes flaccid.

The walnut-shaped **prostate gland** surrounds the urethra, the connection to the urinary bladder. It has a series of short ducts that directly connect to the urethra. The gland is a mixture of smooth muscle and glandular tissue. The muscle provides much of the force needed for ejaculation to occur. The glandular tissue makes a thin, milky fluid that contains citrate (a nutrient), enzymes, and prostate specific antigen (PSA). PSA is a proteolytic enzyme that helps to liquefy the ejaculate several minutes after release from the male. Prostate gland secretions account for about 30 percent of the bulk of semen.

The **bulbourethral gland**, or Cowper’s gland, releases its secretion prior to the release of the bulk of the semen. It neutralizes any acid residue in the urethra left over from urine. This usually accounts for a

couple of drops of fluid in the total ejaculate and may contain a few sperm. Withdrawal of the penis from the vagina before ejaculation to prevent pregnancy may not work if sperm are present in the bulbourethral gland secretions. The location and functions of the male reproductive organs are summarized in Table 24.1.

Table 24.1. Male Reproductive Anatomy

Organ	Location	Function
Scrotum	External	Carry and support testes
Penis	External	Deliver urine, copulating organ
Testes	Internal	Produce sperm and male hormones
Seminal Vesicles	Internal	Contribute to semen production
Prostate Gland	Internal	Contribute to semen production
Bulbourethral Glands	Internal	Clean urethra at ejaculation

Female Reproductive Anatomy

A number of reproductive structures are exterior to the female's body. These include the breasts and the vulva, which consists of the mons pubis, clitoris, labia majora, labia minora, and the vestibular glands, all illustrated in Figure 24.10. The location and functions of the female reproductive organs are summarized in Table 24.2. The vulva is an area associated with the vestibule which includes the structures found in the inguinal (groin) area of women. The mons pubis is a round, fatty area that overlies the pubic symphysis. The **clitoris** is a structure with erectile tissue that contains a large number of sensory nerves and serves as a source of stimulation during intercourse. The **labia majora** are a pair of elongated folds of tissue that run posterior from the mons pubis and enclose the other components of the vulva. The labia majora derive from the same tissue that produces the scrotum in a male. The **labia minora** are thin folds of tissue centrally located within the labia majora. These labia protect the openings to the vagina and urethra. The mons pubis and the anterior portion of the labia majora become covered with hair during adolescence; the labia minora is hairless. The greater vestibular glands are found at the sides of the vaginal opening and provide lubrication during intercourse.

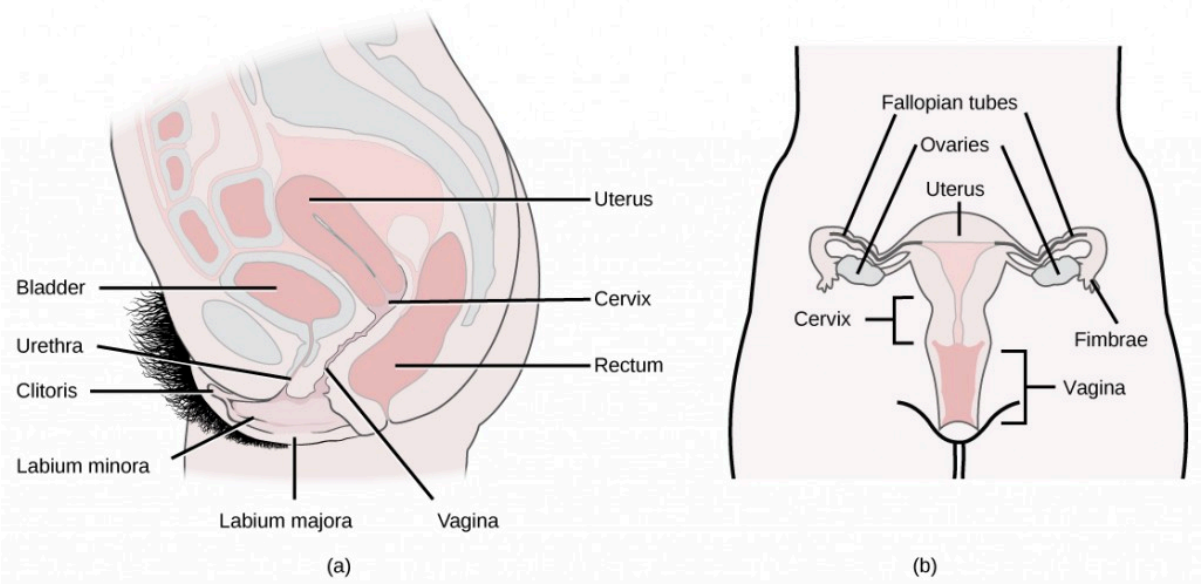


Figure 24.10. The reproductive structures of the human female are shown. (credit a: modification of work by Gray's Anatomy; credit b: modification of work by CDC)

Table 24.2. Female Reproductive Anatomy

Organ	Location	Function
Clitoris	External	Sensory organ
Mons pubis	External	Fatty area overlying pubic bone
Labia majora	External	Covers labia minora
Labia minora	External	Covers vestibule
Greater vestibular glands	External	Secrete mucus; lubricate vagina
Breast	External	Produce and deliver milk
Ovaries	Internal	Carry and develop eggs
Oviducts (Fallopian tubes)	Internal	Transport egg to uterus
Uterus	Internal	Support developing embryo
Vagina	Internal	Common tube for intercourse, birth canal, passing menstrual flow

The breasts consist of mammary glands and fat. The size of the breast is determined by the amount of fat deposited behind the gland. Each gland consists of 15 to 25 lobes that have ducts that empty at the nipple and that supply the nursing child with nutrient- and antibody-rich milk to aid development and protect the child.

Internal female reproductive structures include ovaries, oviducts, the **uterus**, and the vagina, shown in Figure 24.10. The pair of ovaries is held in place in the abdominal cavity by a system of ligaments. Ovaries consist of a medulla and cortex: the medulla contains nerves and blood vessels to supply the cortex with nutrients and remove waste. The outer layers of cells of the cortex are the functional parts

of the ovaries. The cortex is made up of follicular cells that surround eggs that develop during fetal development *in utero*. During the menstrual period, a batch of follicular cells develops and prepares the eggs for release. At ovulation, one follicle ruptures and one egg is released, as illustrated in Figure 24.11**a**.

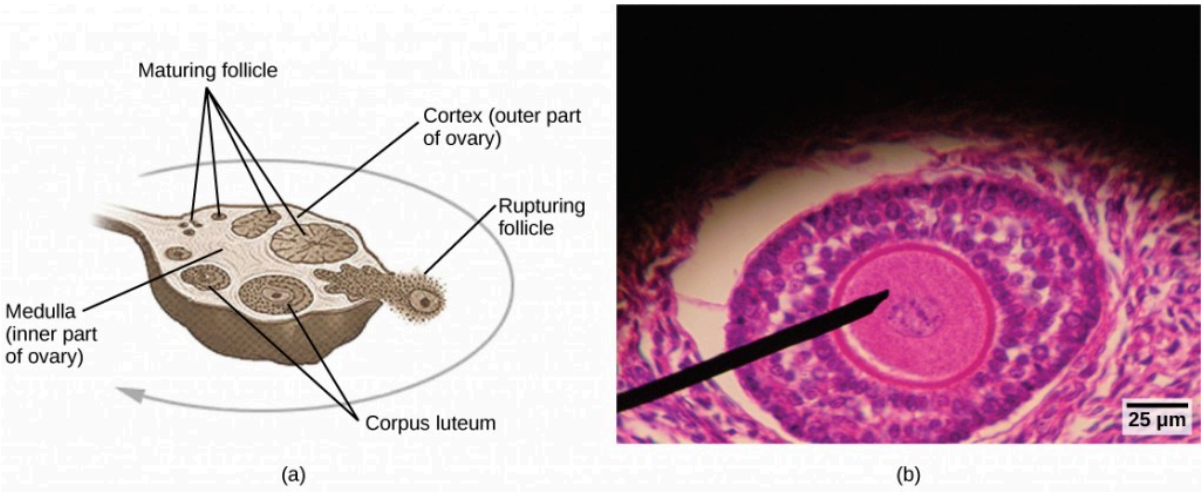


Figure 24.11. Oocytes develop in (a) follicles, located in the ovary. At the beginning of the menstrual cycle, the follicle matures. At ovulation, the follicle ruptures, releasing the egg. The follicle becomes a corpus luteum, which eventually degenerates. The (b) follicle in this light micrograph has an oocyte at its center. (credit a: modification of work by NIH; scale-bar data from Matt Russell)

The **oviducts**, or fallopian tubes, extend from the uterus in the lower abdominal cavity to the ovaries, but they are not in contact with the ovaries. The lateral ends of the oviducts flare out into a trumpet-like structure and have a fringe of finger-like projections called fimbriae, illustrated in Figure 24.10**b**. When an egg is released at ovulation, the fimbrae help the non-motile egg enter into the tube and passage to the uterus. The walls of the oviducts are ciliated and are made up mostly of smooth muscle. The cilia beat toward the middle, and the smooth muscle contracts in the same direction, moving the egg toward the uterus. Fertilization usually takes place within the oviducts and the developing embryo is moved toward the uterus for development. It usually takes the egg or embryo a week to travel through the oviduct. Sterilization in women is called a tubal ligation; it is analogous to a vasectomy in males in that the oviducts are severed and sealed.

The uterus is a structure about the size of a woman’s fist. This is lined with an endometrium rich in blood vessels and mucus glands. The uterus supports the developing embryo and fetus during gestation. The thickest portion of the wall of the uterus is made of smooth muscle. Contractions of the smooth muscle in the uterus aid in passing the baby through the vagina during labor. A portion of the lining of the uterus sloughs off during each menstrual period, and then builds up again in preparation for an implantation. Part of the uterus, called the cervix, protrudes into the top of the vagina. The cervix functions as the birth canal.

The **vagina** is a muscular tube that serves several purposes. It allows menstrual flow to leave the body. It is the receptacle for the penis during intercourse and the vessel for the delivery of offspring. It is lined by stratified squamous epithelial cells to protect the underlying tissue.

Sexual Response during Intercourse

The sexual response in humans is both psychological and physiological. Both sexes experience sexual arousal through psychological and physical stimulation. There are four phases of the sexual response. During phase one, called excitement, vasodilation leads to vasocongestion in erectile tissues in both men and women. The nipples, clitoris, labia, and penis engorge with blood and become enlarged. Vaginal secretions are released to lubricate the vagina to facilitate intercourse. During the second phase, called the plateau, stimulation continues, the outer third of the vaginal wall enlarges with blood, and breathing and heart rate increase.

During phase three, or orgasm, rhythmic, involuntary contractions of muscles occur in both sexes. In the male, the reproductive accessory glands and tubules constrict placing semen in the urethra, then the urethra contracts expelling the semen through the penis. In women, the uterus and vaginal muscles contract in waves that may last slightly less than a second each. During phase four, or resolution, the processes described in the first three phases reverse themselves and return to their normal state. Men experience a refractory period in which they cannot maintain an erection or ejaculate for a period of time ranging from minutes to hours.

Gametogenesis (Spermatogenesis and Oogenesis)

Gametogenesis, the production of sperm and eggs, takes place through the process of meiosis. During meiosis, two cell divisions separate the paired chromosomes in the nucleus and then separate the chromatids that were made during an earlier stage of the cell’s life cycle. Meiosis produces haploid cells with half of each pair of chromosomes normally found in diploid cells. The production of sperm is called **spermatogenesis** and the production of eggs is called **oogenesis**.

Spermatogenesis

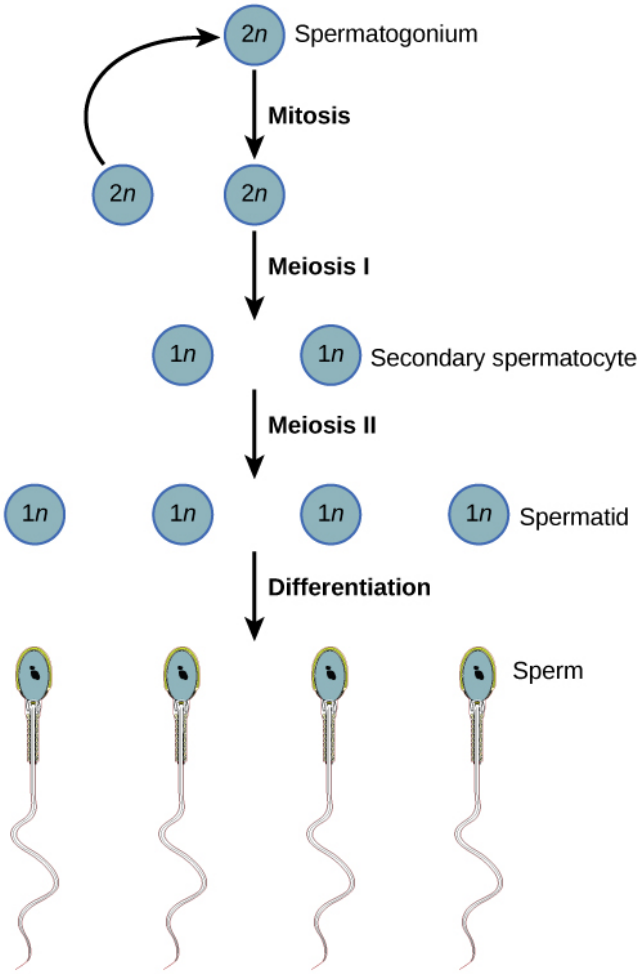


Figure 24.12. During spermatogenesis, four sperm result from each primary spermatocyte.

Spermatogenesis, illustrated in Figure 24.12, occurs in the wall of the seminiferous tubules (Figure 24.8), with stem cells at the periphery of the tube and the spermatozoa at the lumen of the tube. Immediately under the capsule of the tubule are diploid, undifferentiated cells. These stem cells, called spermatogonia (singular: spermatagonium), go through mitosis with one offspring going on to differentiate into a sperm cell and the other giving rise to the next generation of sperm.

Meiosis starts with a cell called a primary spermatocyte. At the end of the first meiotic division, a haploid cell is produced called a secondary spermatocyte. This cell is haploid and must go through another meiotic cell division. The cell produced at the end of meiosis is called a spermatid and when it reaches the lumen of the tubule and grows a flagellum, it is called a sperm cell. Four sperm result from each primary spermatocyte that goes through meiosis.

Stem cells are deposited during gestation and are present at birth through the beginning of adolescence, but in an inactive state. During adolescence, gonadotropic hormones from the anterior pituitary cause the activation of these cells and the production of viable sperm. This continues into old age.

Concept in Action

Visit this site to see the process of spermatogenesis.

Oogenesis

Oogenesis, illustrated in Figure 24.13, occurs in the outermost layers of the ovaries. As with sperm production, oogenesis starts with a germ cell, called an oogonium (plural: oogonia), but this cell undergoes mitosis to increase in number, eventually resulting in up to about one to two million cells in the embryo.

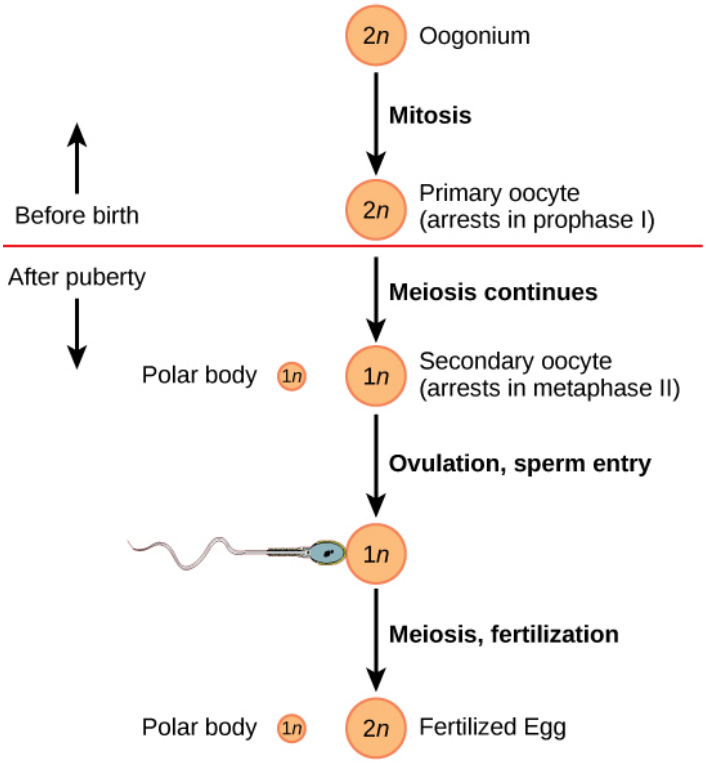


Figure 24.13. The process of oogenesis occurs in the ovary's outermost layer.

The cell starting meiosis is called a primary oocyte, as shown in Figure 24.13. This cell will start the first meiotic division and be arrested in its progress in the first prophase stage. At the time of birth, all future eggs are in the prophase stage. At adolescence, anterior pituitary hormones cause the development of a number of follicles in an ovary. This results in the primary oocyte finishing the first meiotic division. The cell divides unequally, with most of the cellular material and organelles going to

one cell, called a secondary oocyte, and only one set of chromosomes and a small amount of cytoplasm going to the other cell. This second cell is called a polar body and usually dies. A secondary meiotic arrest occurs, this time at the metaphase II stage. At ovulation, this secondary oocyte will be released and travel toward the uterus through the oviduct. If the secondary oocyte is fertilized, the cell continues through the meiosis II, producing a second polar body and a fertilized egg containing all 46 chromosomes of a human being, half of them coming from the sperm.

Egg production begins before birth, is arrested during meiosis until puberty, and then individual cells continue through at each menstrual cycle. One egg is produced from each meiotic process, with the extra chromosomes and chromatids going into polar bodies that degenerate and are reabsorbed by the body.

Summary

As animals became more complex, specific organs and organ systems developed to support specific functions for the organism. The reproductive structures that evolved in land animals allow males and females to mate, fertilize internally, and support the growth and development of offspring. Processes developed to produce reproductive cells that had exactly half the number of chromosomes of each parent so that new combinations would have the appropriate amount of genetic material. Gametogenesis, the production of sperm (spermatogenesis) and eggs (oogenesis), takes place through the process of meiosis.

Exercises

1. Which of the following statements about the male reproductive system is false?
    1. The vas deferens carries sperm from the testes to the penis.
    2. Sperm mature in seminiferous tubules in the testes.
    3. Both the prostate and the bulbourethral glands produce components of the semen.
    4. The prostate gland is located in the testes.
2. Sperm are produced in the _____.
    1. scrotum
    2. seminal vesicles
    3. seminiferous tubules
    4. prostate gland
3. Most of the bulk of semen is made by the _____.
    1. scrotum
    2. seminal vesicles
    3. seminiferous tubules
    4. prostate gland

4. Which of the following cells in spermatogenesis is diploid?
    1. primary spermatocyte
    2. secondary spermatocyte
    3. spermatid
    4. sperm
5. Which female organ has the same embryonic origin as the penis?
    1. clitoris
    2. labia majora
    3. greater vestibular glands
    4. vagina
6. Which female organ has an endometrial lining that will support a developing baby?
    1. labia minora
    2. breast
    3. ovaries
    4. uterus
7. How many eggs are produced as a result of one meiotic series of cell divisions?
    1. one
    2. two
    3. three
    4. four
8. Describe the phases of the human sexual response.
9. Compare spermatogenesis and oogenesis as to timing of the processes and the number and type of cells finally produced.

Answers

1. D
2. C
3. C
4. A
5. A
6. D
7. A
8. In phase one (excitement), vasodilation leads to vasocongestion and enlargement of erectile tissues. Vaginal secretions are released to lubricate the vagina during intercourse. In phase two

(plateau), stimulation continues, the outer third of the vaginal wall enlarges with blood, and breathing and heart rate increase. In phase three (orgasm), rhythmic, involuntary contractions of muscles occur. In the male, reproductive accessory glands and tubules constrict, depositing semen in the urethra; then, the urethra contracts, expelling the semen through the penis. In women, the uterus and vaginal muscles contract in waves that may last slightly less than a second each. In phase four (resolution), the processes listed in the first three phases reverse themselves and return to their normal state. Men experience a refractory period in which they cannot maintain an erection or ejaculate for a period of time ranging from minutes to hours. Women do not experience a refractory period.

9. Stem cells are laid down in the male during gestation and lie dormant until adolescence. Stem cells in the female increase to one to two million and enter the first meiotic division and are arrested in prophase. At adolescence, spermatogenesis begins and continues until death, producing the maximum number of sperm with each meiotic division. Oogenesis continues again at adolescence in batches of oogonia with each menstrual cycle. These oogonia finish the first meiotic division, producing a primary oocyte with most of the cytoplasm and its contents, and a second cell called a polar body containing 23 chromosomes. The second meiotic division results in a secondary oocyte and a second oocyte. At ovulation, a mature haploid egg is released. If this egg is fertilized, it finishes the second meiotic division, including the chromosomes donated by the sperm in the finished cell. This is a diploid, fertilized egg.

Glossary

bulbourethral gland
secretion that cleanses the urethra prior to ejaculation

clitoris
sensory structure in females; stimulated during sexual arousal

labia majora
large folds of tissue covering the inguinal area

labia minora
smaller folds of tissue within the labia majora

oogenesis
process of producing haploid eggs

oviduct(also, fallopian tube)
muscular tube connecting the uterus with the ovary area

penismale reproductive
structure for urine elimination and copulation

prostate gland
structure that is a mixture of smooth muscle and glandular material and that contributes to semen

scrotum
sac containing testes; exterior to the body

semen
fluid mixture of sperm and supporting materials

seminal vesicle
secretory accessory gland in males; contributes to semen

seminiferous tubule
site of sperm production in testes

spermatogenesis
process of producing haploid sperm

testes
pair of reproductive organs in males

uterus
environment for developing embryo and fetus

vagina
muscular tube for the passage of menstrual flow, copulation, and birth of offspring

24.4. Hormonal Control of Human Reproduction

Learning Objectives

By the end of this chapter, you will be able to:

- Describe the roles of male and female reproductive hormones
- Discuss the interplay of the ovarian and menstrual cycles
- Describe the process of menopause

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 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. Note that the body must reach puberty in order for the adrenals to release the hormones that must be present for GnRH to be produced. Although FSH and LH 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** to begin facilitating spermatogenesis using negative feedback, as illustrated in

Figure 24.14. LH also enters the testes and stimulates the **interstitial cells of Leydig** to make and release testosterone into the testes and the blood.

Testosterone, the hormone responsible for the secondary sexual characteristics that develop in the male during adolescence, stimulates spermatogenesis. These secondary sex characteristics include a deepening of the voice, the growth of facial, axillary, and pubic hair, and the beginnings of the sex drive.

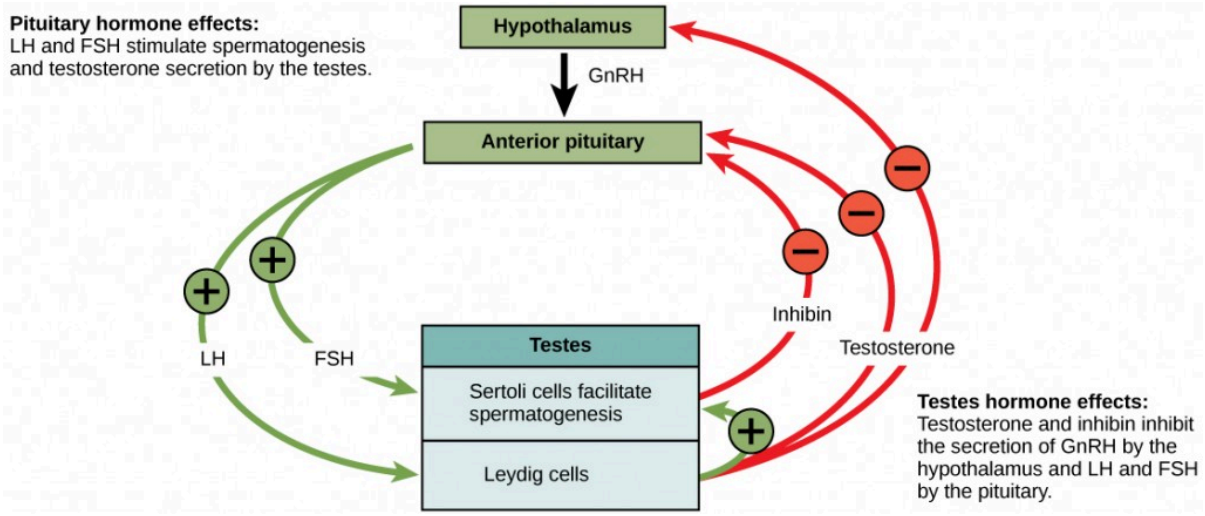


Figure 24.14. Hormones control sperm production in a negative feedback system.

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. 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 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. As with the male, the anterior pituitary hormones cause the release of the hormones FSH and LH. In addition, estrogens and progesterone are released from the developing follicles. **Estrogen** is the reproductive hormone in females that assists in endometrial regrowth, ovulation, and calcium absorption; it is also responsible for the secondary sexual characteristics of females. These include breast development, flaring of the hips, and a shorter period necessary for bone maturation. **Progesterone** assists in endometrial re-growth and inhibition of FSH and LH release.

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. 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.

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. These cycles occur concurrently and are coordinated over a 22–32 day cycle, with an average length of 28 days.

The first half of the ovarian cycle is the follicular phase shown in Figure 24.15. Slowly rising levels of FSH and LH 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 estrogens and a low level of progesterone. Progesterone maintains the endometrium to help ensure pregnancy. The trip through the fallopian tube takes about seven days. At this stage of development, called the morula, there are 30-60 cells. If pregnancy implantation does not occur, the lining is sloughed off. 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.

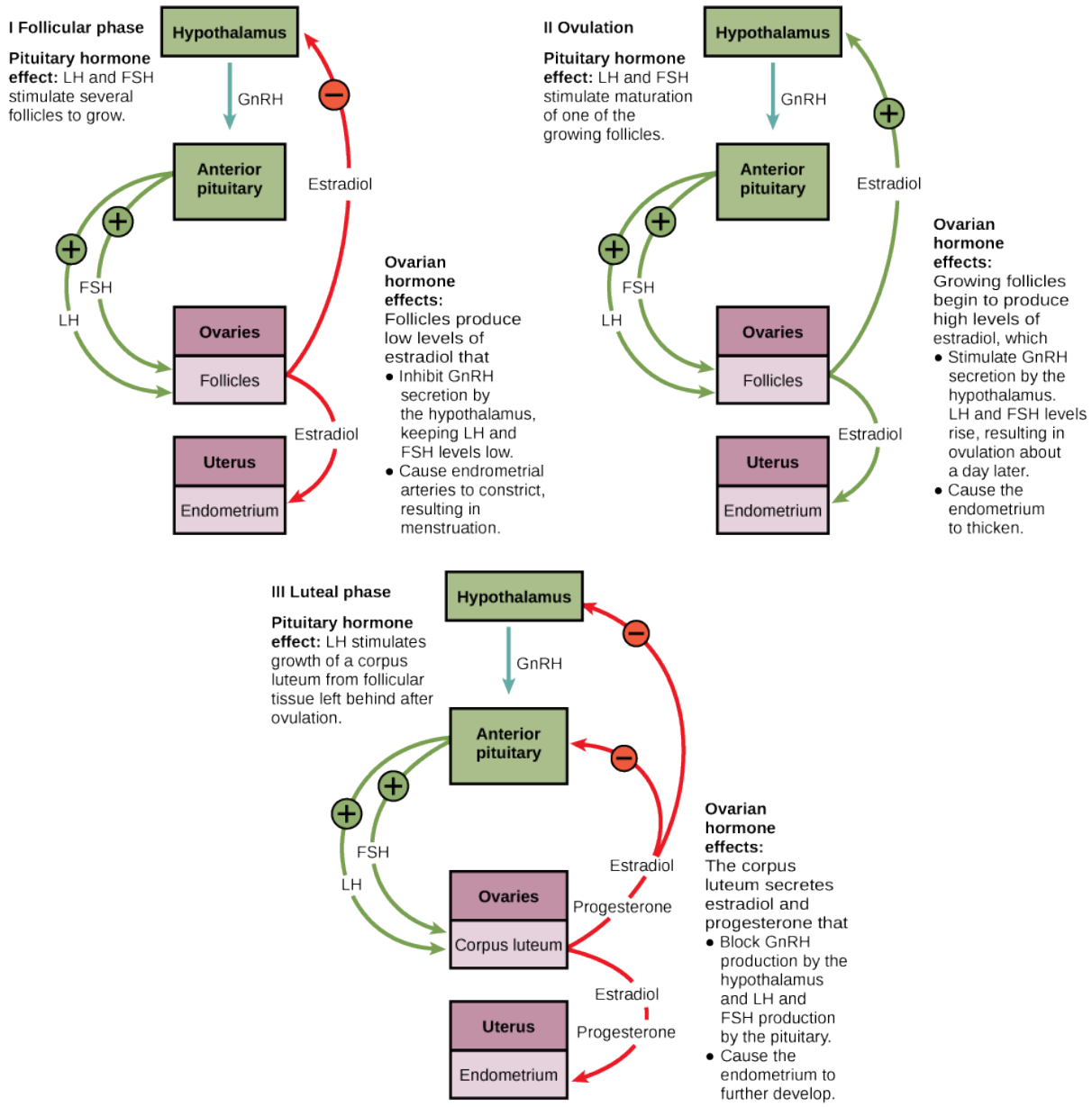


Figure 24.15. The ovarian and menstrual cycles of female reproduction are regulated by hormones produced by the hypothalamus, pituitary, and ovaries.

Which of the following statements about hormone regulation of the female reproductive cycle is false?

1. LH and FSH are produced in the pituitary, and estradiol and progesterone are produced in the

ovaries.

2. Estradiol and progesterone secreted from the corpus luteum cause the endometrium to thicken.
3. Both progesterone and estradiol are produced by the follicles.
4. Secretion of GnRH by the hypothalamus is inhibited by low levels of estradiol but stimulated by high levels of estradiol.

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 **ovulation**: the most mature follicle, like that shown in Figure 24.16, ruptures and releases its egg. The follicles that did not rupture degenerate and their eggs are lost. The level of estrogen decreases when the extra follicles degenerate.

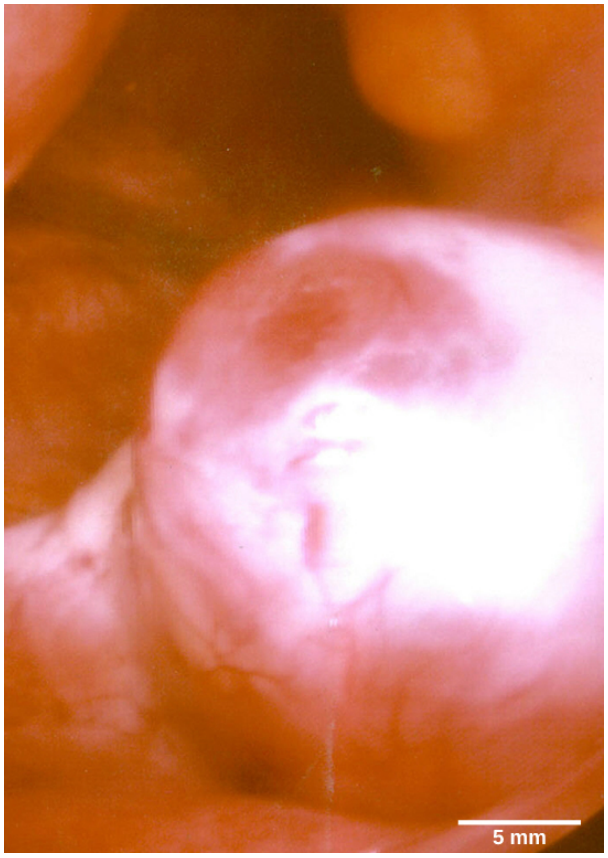


Figure 24.16. This mature egg follicle may rupture and release an egg. (credit: scale-bar data from Matt Russell)

Following ovulation, the ovarian cycle enters its luteal phase, illustrated in Figure 24.15 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. Figure 24.17 visually compares the ovarian and uterine cycles as well as the commensurate hormone levels.

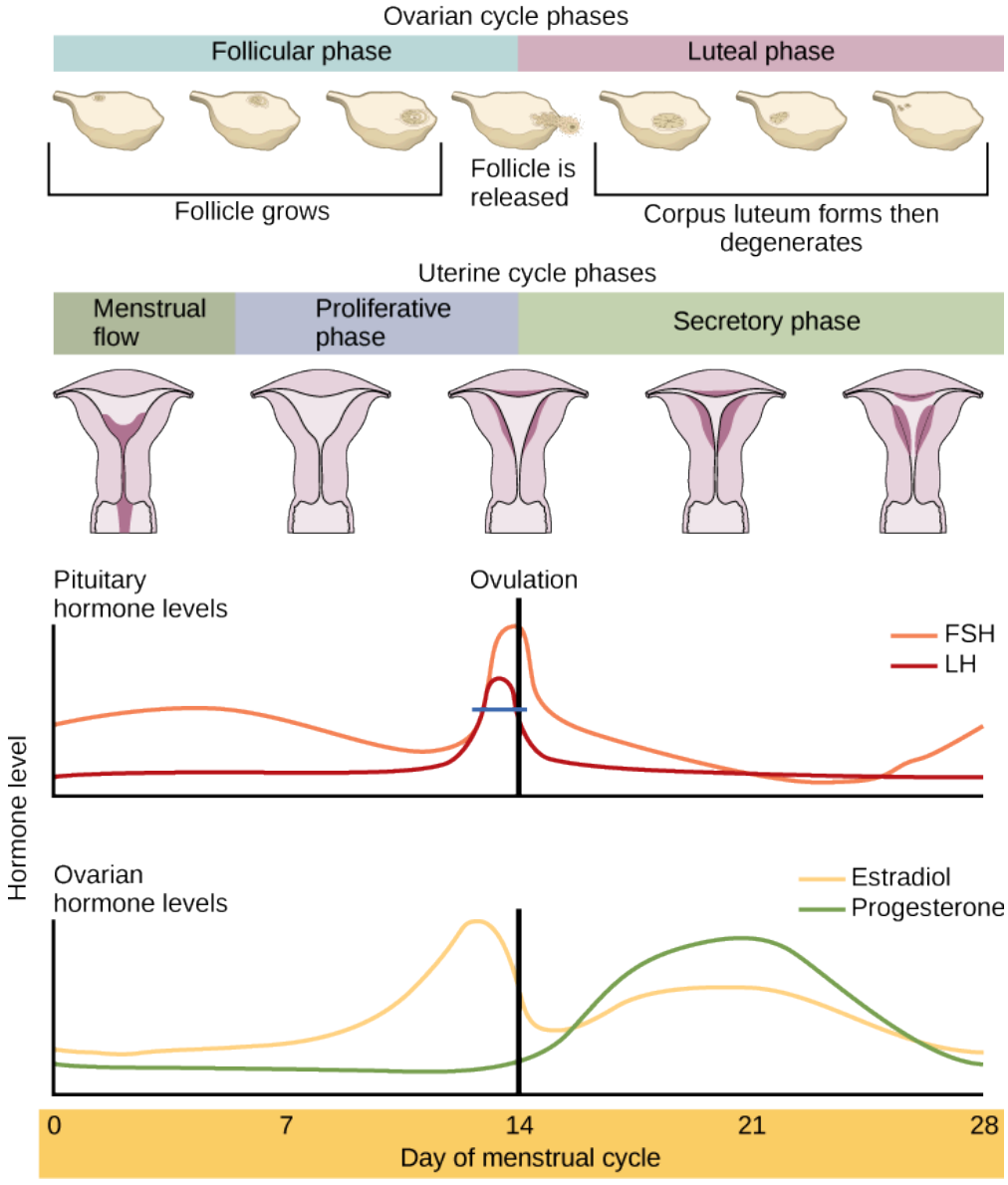


Figure 24.17. Rising and falling hormone levels result in progression of the ovarian and menstrual cycles. (credit: modification of work by Mikael Häggström)

Which of the following statements about the menstrual cycle is false?

1. Progesterone levels rise during the luteal phase of the ovarian cycle and the secretory phase of the uterine cycle.

2. Menstruation occurs just after LH and FSH levels peak.
3. Menstruation occurs after progesterone levels drop.
4. Estrogen levels rise before ovulation, while progesterone levels rise after.

Menopause

As women approach their mid-40s to mid-50s, their ovaries begin to lose their sensitivity to FSH and LH. Menstrual periods become less frequent and finally cease; this is **menopause**. There are still eggs and potential follicles on the ovaries, but without the stimulation of FSH and LH, they will not produce a viable egg to be released. The outcome of this is the inability to have children.

The side effects of menopause include hot flashes, heavy sweating (especially at night), headaches, some hair loss, muscle pain, vaginal dryness, insomnia, depression, weight gain, and mood swings. Estrogen is involved in calcium metabolism and, without it, blood levels of calcium decrease. To replenish the blood, calcium is lost from bone which may decrease the bone density and lead to osteoporosis. Supplementation of estrogen in the form of hormone replacement therapy (HRT) can prevent bone loss, but the therapy can have negative side effects. While HRT is thought to give some protection from colon cancer, osteoporosis, heart disease, macular degeneration, and possibly depression, its negative side effects include increased risk of: stroke or heart attack, blood clots, breast cancer, ovarian cancer, endometrial cancer, gall bladder disease, and possibly dementia.

Reproductive Endocrinologist

A reproductive endocrinologist is a physician who treats a variety of hormonal disorders related to reproduction and infertility in both men and women. The disorders include menstrual problems, infertility, pregnancy loss, sexual dysfunction, and menopause. Doctors may use fertility drugs, surgery, or assisted reproductive techniques (ART) in their therapy. ART involves the use of procedures to manipulate the egg or sperm to facilitate reproduction, such as *in vitro* fertilization.

Reproductive endocrinologists undergo extensive medical training, first in a four-year residency in obstetrics and gynecology, then in a three-year fellowship in reproductive endocrinology. To be board certified in this area, the physician must pass written and oral exams in both areas.

Summary

The male and female reproductive cycles are controlled by hormones released from the hypothalamus and anterior pituitary as well as hormones from reproductive tissues and organs. The hypothalamus monitors the need for the FSH and LH hormones made and released from the anterior pituitary. FSH and LH affect reproductive structures to cause the formation of sperm and the preparation of eggs for release and possible fertilization. In the male, FSH and LH stimulate Sertoli cells and interstitial cells of Leydig in the testes to facilitate sperm production. The Leydig cells produce testosterone, which also is responsible for the secondary sexual characteristics of males. In females, FSH and LH cause estrogen and progesterone to be produced. They regulate the female reproductive system which is divided into the ovarian cycle and the menstrual cycle. Menopause occurs when the ovaries lose their sensitivity to FSH and LH and the female reproductive cycles slow to a stop.

Exercises

1. Which of the following statements about hormone regulation of the female reproductive cycle is false?
    1. LH and FSH are produced in the pituitary, and estradiol and progesterone are produced in the ovaries.
    2. Estradiol and progesterone secreted from the corpus luteum cause the endometrium to thicken.
    3. Both progesterone and estradiol are produced by the follicles.
    4. Secretion of GnRH by the hypothalamus is inhibited by low levels of estradiol but stimulated by high levels of estradiol.
2. Which of the following statements about the menstrual cycle is false?
    1. Progesterone levels rise during the luteal phase of the ovarian cycle and the secretory phase of the uterine cycle.
    2. Menstruation occurs just after LH and FSH levels peak.
    3. Menstruation occurs after progesterone levels drop.
    4. Estrogen levels rise before ovulation, while progesterone levels rise after.
3. Which hormone causes Leydig cells to make testosterone?
    1. FSH
    2. LH
    3. inhibin
    4. estrogen
4. Which hormone causes FSH and LH to be released?
    1. testosterone
    2. estrogen
    3. GnRH
    4. progesterone
5. Which hormone signals ovulation?
    1. FSH
    2. LH
    3. inhibin
    4. estrogen
6. Which hormone causes the re-growth of the endometrial lining of the uterus?
    1. testosterone

2. estrogen
3. GnRH
4. progesterone

7. If male reproductive pathways are not cyclical, how are they controlled?
8. Describe the events in the ovarian cycle leading up to ovulation.

Answers

1. C
2. B
3. A
4. C
5. B
6. D
7. Negative feedback in the male system is supplied through two hormones: inhibin and testosterone. Inhibin is produced by Sertoli cells when the sperm count exceeds set limits. The hormone inhibits GnRH and FSH, decreasing the activity of the Sertoli cells. Increased levels of testosterone affect the release of both GnRH and LH, decreasing the activity of the Leydig cells, resulting in decreased testosterone and sperm production.
8. Low levels of progesterone allow the hypothalamus to send GnRH to the anterior pituitary and cause the release of FSH and LH. FSH stimulates follicles on the ovary to grow and prepare the eggs for ovulation. As the follicles increase in size, they begin to release estrogen and a low level of progesterone into the blood. The level of estrogen rises to a peak, causing a spike in the concentration of LH. This causes the most mature follicle to rupture and ovulation occurs.

Glossary

estrogen
reproductive hormone in females that assists in endometrial regrowth, ovulation, and calcium absorption

follicle stimulating hormone (FSH)
reproductive hormone that causes sperm production in men and follicle development in women

gonadotropin-releasing hormone (GnRH)
hormone from the hypothalamus that causes the release of FSH and LH from the anterior pituitary

inhibin
hormone made by Sertoli cells; provides negative feedback to hypothalamus in control of FSH and GnRH release

interstitial
cell of Leydigcell in seminiferous tubules that makes testosterone

luteinizing hormone (LH)

reproductive hormone in both men and women, causes testosterone production in men and ovulation and lactation in women

menopause
loss of reproductive capacity in women due to decreased sensitivity of the ovaries to FSH and LH

menstrual cycle
cycle of the degradation and re-growth of the endometrium

ovarian cycle
cycle of preparation of egg for ovulation and the conversion of the follicle to the corpus luteum

ovulation
release of the egg by the most mature follicle

progesterone
reproductive hormone in women; assists in endometrial re-growth and inhibition of FSH and LH release

Sertoli cell
cell in seminiferous tubules that assists developing sperm and makes inhibin

testosterone
reproductive hormone in men that assists in sperm production and promoting secondary sexual characteristics

24.5. Human Pregnancy and Birth

Learning Objectives

By the end of this section, you will be able to:

- Explain fetal development during the three trimesters of gestation
- Describe labor and delivery
- Compare the efficacy and duration of various types of contraception
- Discuss causes of infertility and the therapeutic options available

Pregnancy begins with the fertilization of an egg and continues through to the birth of the individual. The length of time of **gestation** varies among animals, but is very similar among the great apes: human gestation is 266 days, while chimpanzee gestation is 237 days, a gorilla's is 257 days, and orangutan gestation is 260 days long. The fox has a 57-day gestation. Dogs and cats have similar gestations averaging 60 days. The longest gestation for a land mammal is an African elephant at 640 days. The longest gestations among marine mammals are the beluga and sperm whales at 460 days.

Human Gestation

Twenty-four hours before fertilization, the egg has finished meiosis and becomes a mature oocyte. When fertilized (at conception) the egg becomes known as a zygote. The zygote travels through the oviduct to the uterus (Figure 24.18). 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 zygote (blastocyst) grow into the endometrium by digesting the endometrial cells, and wound 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.

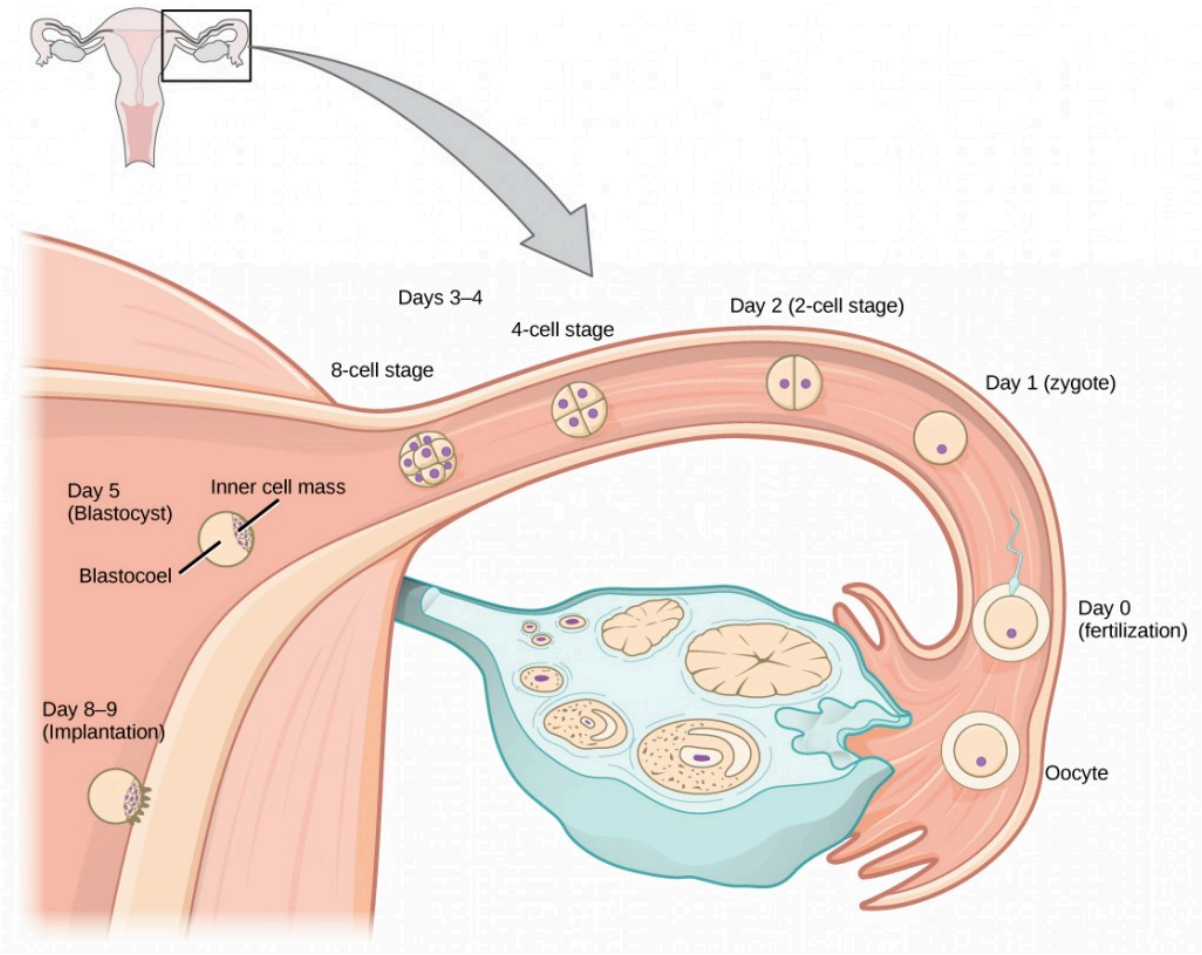


Figure 24.18. In humans, fertilization occurs soon after the oocyte leaves the ovary. Implantation occurs eight or nine days later.(credit: Ed Uthman)

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. This organ 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, as shown in Figure 24.19. 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 affects that development can have a severe effect on the fetus' survival.

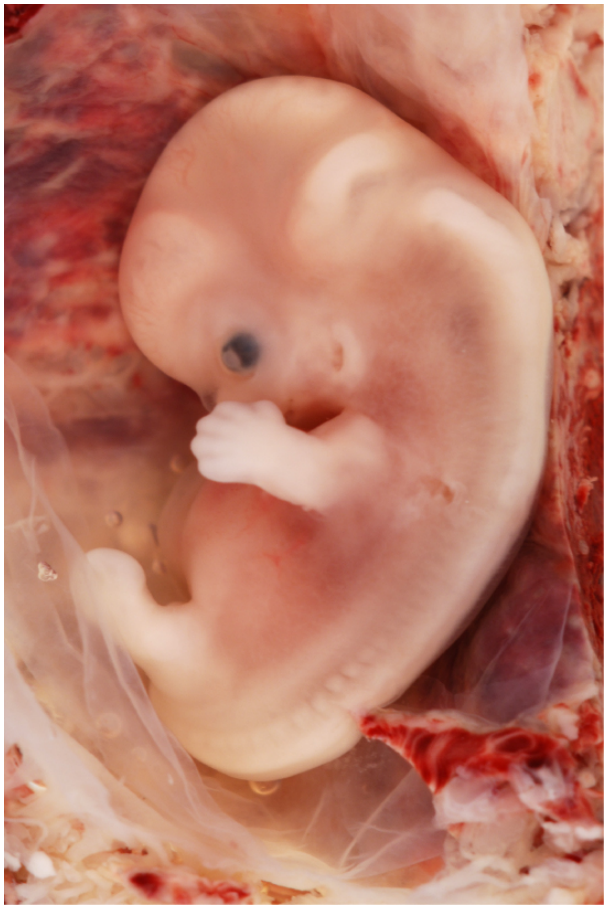


Figure 24.19. Fetal development is shown at nine weeks gestation. (credit: Ed Uthman)

During the second trimester, the fetus grows to about 30 cm (12 inches), as shown in Figure 24.20. 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 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.

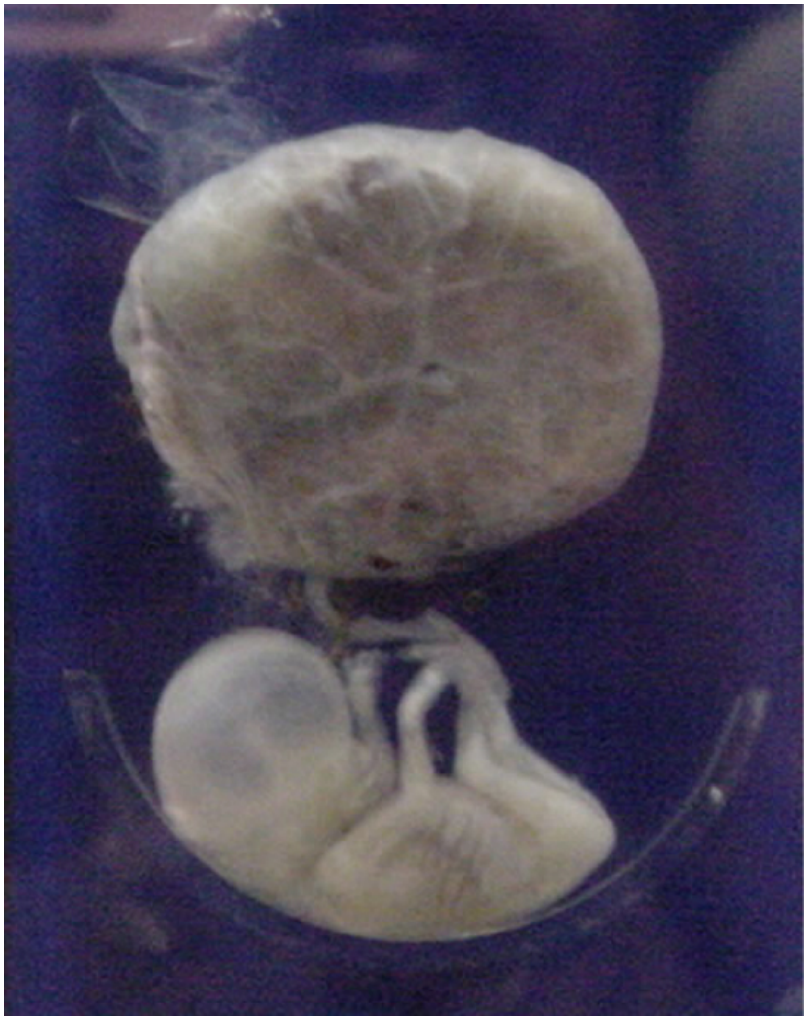


Figure 24.20. This fetus is just entering the second trimester, when the placenta takes over more of the functions performed as the baby develops. (credit: National Museum of Health and Medicine)

During the third trimester, the fetus grows to 3 to 4 kg (6 ½ -8 ½ lbs.) and about 50 cm (19-20 inches) long, as illustrated in Figure 24.21. This is the period of the most rapid growth during the pregnancy. Organ development continues to birth (and some systems, such as the nervous system and liver, continue to develop after birth). The mother will be at her most uncomfortable during this trimester. She may urinate frequently due to pressure on the bladder from the fetus. There may also be intestinal blockage and circulatory problems, especially in her legs. Clots may form in her legs due to pressure from the fetus on returning veins as they enter the abdominal cavity.

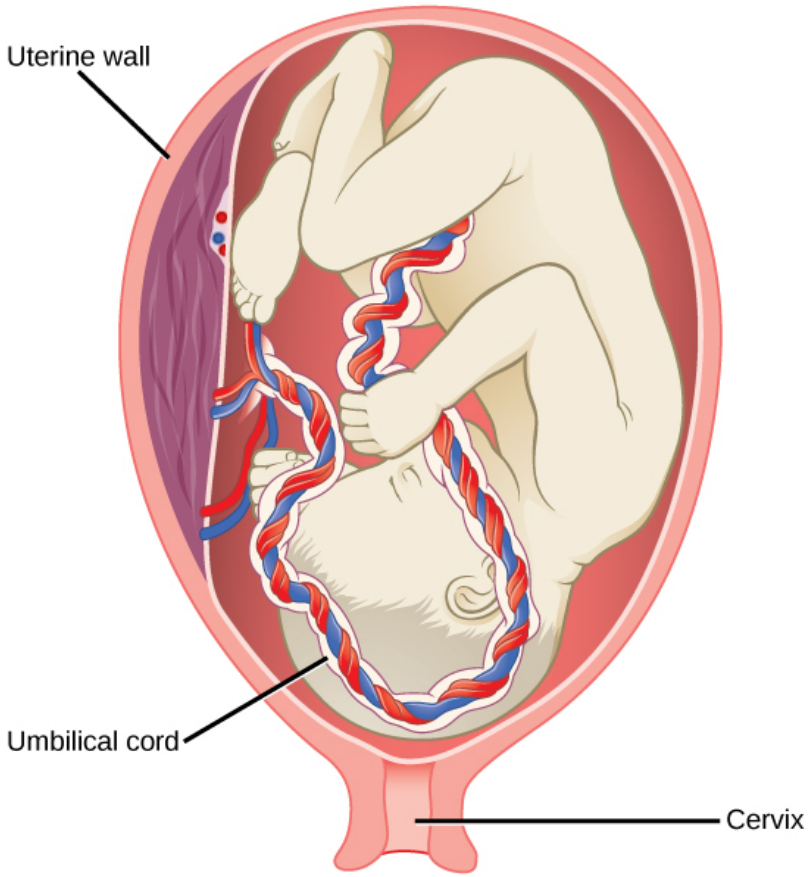


Figure 24.21. There is rapid fetal growth during the third trimester. (credit: modification of work by Gray's Anatomy)

Concept in Action

Visit this site to see the stages of human fetal development.

Labor and Birth

Labor is the physical efforts of expulsion of the fetus and the placenta from the uterus during birth (parturition). 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 baby reorients, facing forward and down with the back or crown of the head engaging the cervix (uterine opening). This causes the cervix to stretch and nerve impulses are sent to the hypothalamus, which signals for the release of oxytocin from the

posterior pituitary. The oxytocin causes the 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.

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.

An alternative to labor and delivery is the surgical delivery of the baby through a procedure called a Caesarian section. This is major abdominal surgery and can lead to post-surgical complications for the mother, but in some cases it may be the only way to safely deliver the baby.

The mother’s mammary glands go through changes during the third trimester to prepare for lactation and breastfeeding. When the baby begins suckling at the breast, signals are sent to the hypothalamus causing the release of prolactin from the anterior pituitary. Prolactin causes the mammary glands to produce milk. Oxytocin is also released, promoting the release of the milk. The milk contains nutrients for the baby’s development and growth as well as immunoglobulins to protect the child from bacterial and viral infections.

Contraception and Birth Control

The prevention of a pregnancy comes under the terms contraception or birth control. Strictly speaking, **contraception** refers to preventing the sperm and egg from joining. Both terms are, however, frequently used interchangeably.

Table 24.3. Contraceptive Methods

Method	Examples	Failure Rate in Typical Use Over 12 Months
Barrier	male condom, female condom, sponge, cervical cap, diaphragm, spermicides	15 to 24%
Hormonal	oral, patch, vaginal ring	8%
	injection	3%
	implant	less than 1%
Other	natural family planning	12 to 25%
	withdrawal	27%
	sterilization	less than 1%

Table 24.3 lists common methods of contraception. The failure rates listed are not the ideal rates that could be realized, but the typical rates that occur. A failure rate is the number of pregnancies resulting from the method's use over a twelve-month period. Barrier methods, such as condoms, cervical caps, and diaphragms, block sperm from entering the uterus, preventing fertilization. Spermicides are chemicals that are placed in the vagina that kill sperm. Sponges, which are saturated with spermicides, are placed in the vagina at the cervical opening. Combinations of spermicidal chemicals and barrier methods achieve lower failure rates than do the methods when used separately.

Nearly a quarter of the couples using barrier methods, natural family planning, or withdrawal can expect a failure of the method. Natural family planning is based on the monitoring of the menstrual cycle and having intercourse only during times when the egg is not available. A woman's body temperature may rise a degree Celsius at ovulation and the cervical mucus may increase in volume and become more pliable. These changes give a general indication of when intercourse is more or less likely to result in fertilization. Withdrawal involves the removal of the penis from the vagina during intercourse, before ejaculation occurs. This is a risky method with a high failure rate due to the possible presence of sperm in the bulbourethral gland's secretion, which may enter the vagina prior to removing the penis.

Hormonal methods use synthetic progesterone (sometimes in combination with estrogen), to inhibit the hypothalamus from releasing FSH or LH, and thus prevent an egg from being available for fertilization. The method of administering the hormone affects failure rate. The most reliable method, with a failure rate of less than 1 percent, is the implantation of the hormone under the skin. The same rate can be achieved through the sterilization procedures of vasectomy in the man or of tubal ligation in the woman, or by using an intrauterine device (IUD). IUDs are inserted into the uterus and establish an inflammatory condition that prevents fertilized eggs from implanting into the uterine wall.

Compliance with the contraceptive method is a strong contributor to the success or failure rate of any particular method. The only method that is completely effective at preventing conception is abstinence. The choice of contraceptive method depends on the goals of the woman or couple. Tubal ligation and vasectomy are considered permanent prevention, while other methods are reversible and provide short-term contraception.

Termination of an existing pregnancy can be spontaneous or voluntary. Spontaneous termination is a miscarriage and usually occurs very early in the pregnancy, usually within the first few weeks. This occurs when the fetus cannot develop properly and the gestation is naturally terminated. Voluntary termination of a pregnancy is an abortion. Laws regulating abortion vary between states and tend to view fetal viability as the criteria for allowing or preventing the procedure.

Infertility

Infertility is the inability to conceive a child or carry a child to birth. About 75 percent of causes of infertility can be identified; these include diseases, such as sexually transmitted diseases that can cause scarring of the reproductive tubes in either men or women, or developmental problems frequently related to abnormal hormone levels in one of the individuals. Inadequate nutrition, especially starvation, can delay menstruation. Stress can also lead to infertility. Short-term stress can affect hormone levels, while long-term stress can delay puberty and cause less frequent menstrual cycles.

Other factors that affect fertility include toxins (such as cadmium), tobacco smoking, marijuana use, gonadal injuries, and aging.

If infertility is identified, several assisted reproductive technologies (ART) are available to aid conception. A common type of ART is *in vitro* fertilization (IVF) where an egg and sperm are combined outside the body and then placed in the uterus. Eggs are obtained from the woman after extensive hormonal treatments that prepare mature eggs for fertilization and prepare the uterus for implantation of the fertilized egg. Sperm are obtained from the man and they are combined with the eggs and supported through several cell divisions to ensure viability of the zygotes. When the embryos have reached the eight-cell stage, one or more is implanted into the woman's uterus. If fertilization is not accomplished by simple IVF, a procedure that injects the sperm into an egg can be used. This is called intracytoplasmic sperm injection (ICSI) and is shown in Figure 24.22. IVF procedures produce a surplus of fertilized eggs and embryos that can be frozen and stored for future use. The procedures can also result in multiple births.

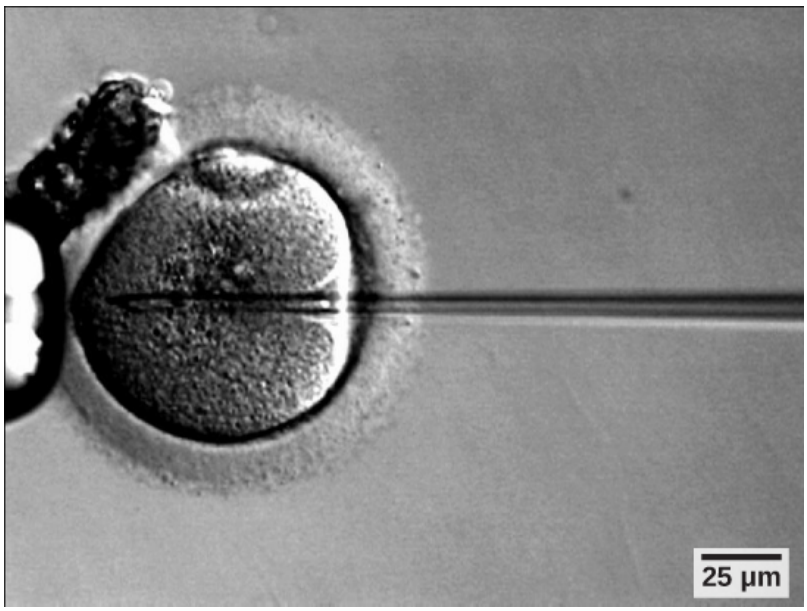


Figure 24.22. A sperm is inserted into an egg for fertilization during intracytoplasmic sperm injection (ICSI). (credit: scale-bar data from Matt Russell)

Summary

Human pregnancy begins with fertilization of an egg and proceeds through the three trimesters of gestation. The labor process has three stages (contractions, delivery of the fetus, expulsion of the placenta), each propelled by hormones. The first trimester lays down the basic structures of the body, including the limb buds, heart, eyes, and the liver. The second trimester continues the development of all of the organs and systems. The third trimester exhibits the greatest growth of the fetus and culminates in labor and delivery. Prevention of a pregnancy can be accomplished through a variety of methods including barriers, hormones, or other means. Assisted reproductive technologies may help individuals who have infertility problems.

Exercises

1. Nutrient and waste requirements for the developing fetus are handled during the first few weeks by:
    1. the placenta
    2. diffusion through the endometrium
    3. the chorion
    4. the blastocyst
2. Progesterone is made during the third trimester by the:
    1. placenta
    2. endometrial lining
    3. chorion
    4. corpus luteum
3. Which contraceptive method is 100 percent effective at preventing pregnancy?
    1. condom
    2. oral hormonal methods
    3. sterilization
    4. abstinence
4. Which type of short term contraceptive method is generally more effective than others?
    1. barrier
    2. horomonal
    3. natural family planning
    4. withdrawal
5. Which hormone is primarily responsible for the contractions during labor?
    1. oxytocin
    2. estrogen
    3. β-HCG
    4. progesterone
6. Major organs begin to develop during which part of human gestation?
    1. fertilization
    2. first trimester
    3. second trimester
    4. third trimester

7. Describe the major developments during each trimester of human gestation.
8. Describe the stages of labor.

Answers

1. B
2. A
3. D
4. B
5. A
6. B
7. The first trimester lays down the basic structures of the body, including the limb buds, heart, eyes, and the liver. The second trimester continues the development of all of the organs and systems established during the first trimester. The placenta takes over the production of estrogen and high levels of progesterone and handles the nutrient and waste requirements of the fetus. The third trimester exhibits the greatest growth of the fetus, culminating in labor and delivery.
8. Stage one of labor results in the thinning of the cervix and the dilation of the cervical opening. Stage two delivers the baby, and stage three delivers the placenta.

Glossary

contraception (also, birth control)
various means used to prevent pregnancy

gestation
length of time for fetal development to birth

human beta chorionic gonadotropin (*β*-HCG)
hormone produced by the chorion of the zygote that helps to maintain the corpus luteum and elevated levels of progesterone

infertility
inability to conceive, carry, and deliver children

placenta
organ that supports the diffusion of nutrients and waste between the mother’s and fetus’ blood

24.6. Fertilization and Early Embryonic Development

Learning Objectives

By the end of this section, you will be able to:

- Discuss how fertilization occurs
- Explain how the embryo forms from the zygote
- Discuss the role of cleavage and gastrulation in animal development

The process in which an organism develops from a single-celled zygote to a multi-cellular organism is complex and well-regulated. The early stages of embryonic development are also crucial for ensuring the fitness of the organism.

Fertilization

Fertilization, pictured in Figure 24.23**a** is the process in which gametes (an egg and sperm) fuse to form a zygote. The egg and sperm each contain one set of chromosomes. To ensure that the offspring has only one complete diploid set of chromosomes, only one sperm must fuse with one egg. In mammals, the egg is protected by a layer of extracellular matrix consisting mainly of glycoproteins called the **zona pellucida**. When a sperm binds to the zona pellucida, a series of biochemical events, called the **acrosomal reactions**, take place. In placental mammals, the acrosome contains digestive enzymes that initiate the degradation of the glycoprotein matrix protecting the egg and allowing the sperm plasma membrane to fuse with the egg plasma membrane, as illustrated in Figure 24.23**b**. The fusion of these two membranes creates an opening through which the sperm nucleus is transferred into the ovum. The nuclear membranes of the egg and sperm break down and the two haploid genomes condense to form a diploid genome.

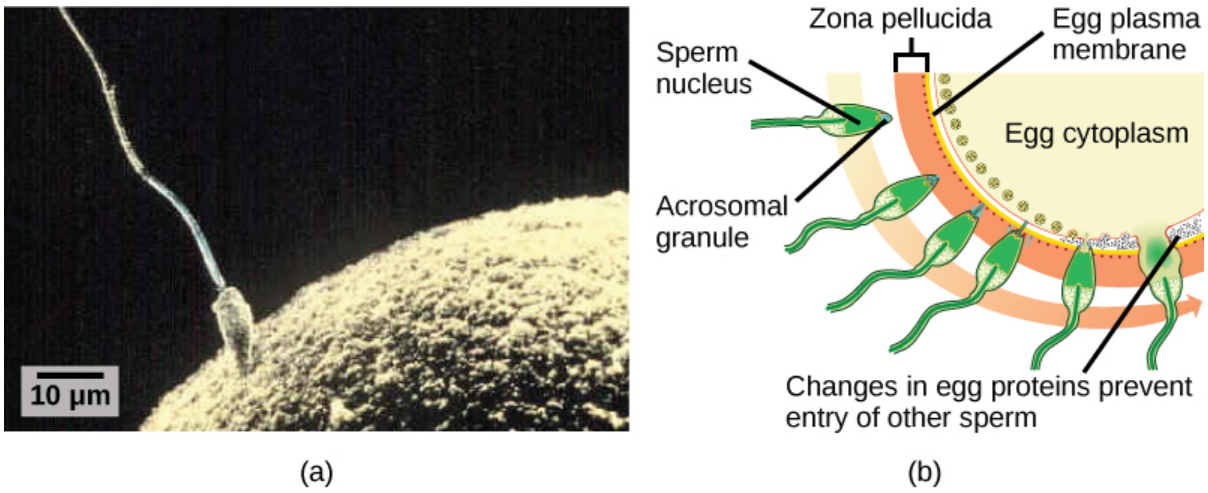


Figure 24.23. (a) Fertilization is the process in which sperm and egg fuse to form a zygote. (b) Acrosomal reactions help the sperm degrade the glycoprotein matrix protecting the egg and allow the sperm to transfer its nucleus. (credit: (b) modification of work by Mariana Ruiz Villareal; scale-bar data from Matt Russell)

To ensure that no more than one sperm fertilizes the egg, once the acrosomal reactions take place at one location of the egg membrane, the egg releases proteins in other locations to prevent other sperm from fusing with the egg. If this mechanism fails, multiple sperm can fuse with the egg, resulting in **polyspermy**. The resulting embryo is not genetically viable and dies within a few days.

Cleavage and Blastula Stage

The development of multi-cellular organisms begins from a single-celled zygote, which undergoes rapid cell division to form the blastula. The rapid, multiple rounds of cell division are termed cleavage. Cleavage is illustrated in (Figure 24.24**a**). After the cleavage has produced over 100 cells, the embryo is called a blastula. The blastula is usually a spherical layer of cells (the blastoderm) surrounding a fluid-filled or yolk-filled cavity (the blastocoel). Mammals at this stage form a structure called the blastocyst, characterized by an inner cell mass that is distinct from the surrounding blastula, shown in Figure 24.24**b**. During cleavage, the cells divide without an increase in mass; that is, one large single-celled zygote divides into multiple smaller cells. Each cell within the blastula is called a blastomere.

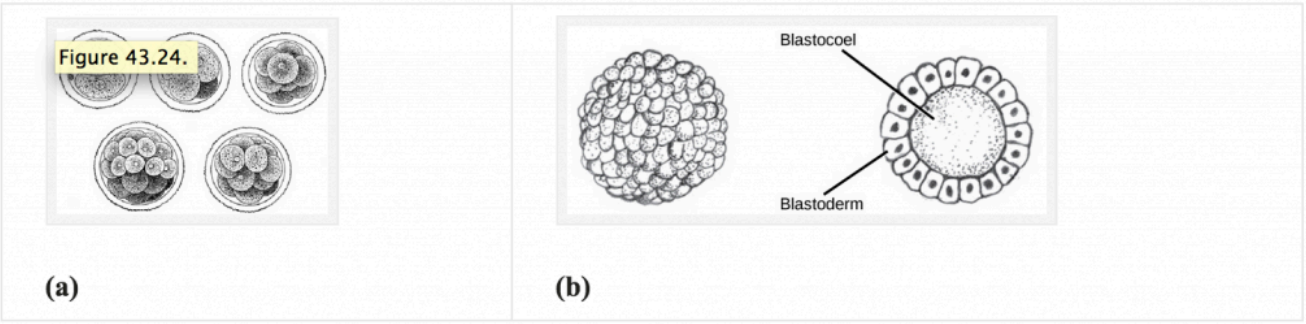


Figure 24.24. (a) During cleavage, the zygote rapidly divides into multiple cells without increasing in size. (b) The cells rearrange themselves to form a hollow ball with a fluid-filled or yolk-filled cavity called the blastula. (credit a: modification of work by Gray's Anatomy; credit b: modification of work by Pearson Scott Foresman, donated to the Wikimedia Foundation)

Cleavage can take place in two ways: **holoblastic** (total) cleavage or **meroblastic** (partial) cleavage. The type of cleavage depends on the amount of yolk in the eggs. In placental mammals (including humans) where nourishment is provided by the mother's body, the eggs have a very small amount of yolk and undergo holoblastic cleavage. Other species, such as birds, with a lot of yolk in the egg to nourish the embryo during development, undergo meroblastic cleavage.

In mammals, the blastula forms the **blastocyst** in the next stage of development. Here the cells in the blastula arrange themselves in two layers: the **inner cell mass**, and an outer layer called the **trophoblast**. The inner cell mass is also known as the embryoblast and this mass of cells will go on to form the embryo. At this stage of development, illustrated in Figure 24.25 the inner cell mass consists of embryonic stem cells that will differentiate into the different cell types needed by the organism. The trophoblast will contribute to the placenta and nourish the embryo.

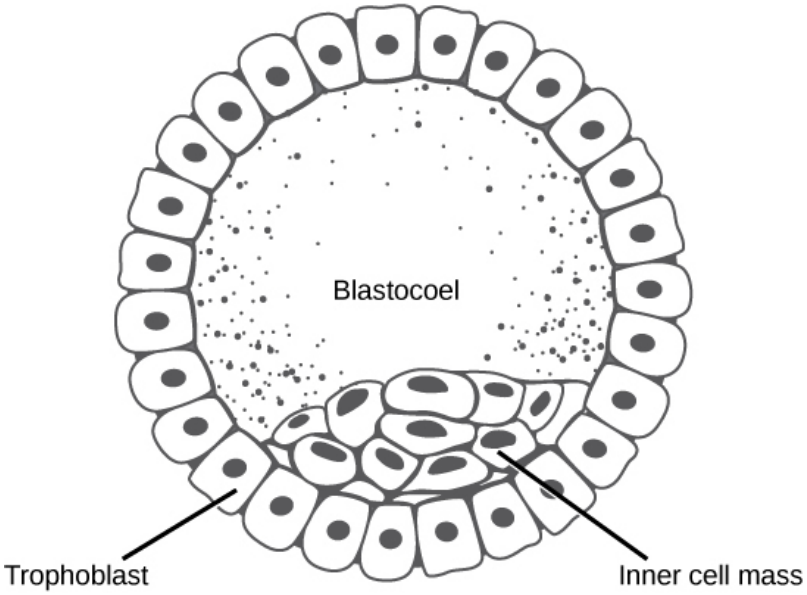


Figure 24.25. The rearrangement of the cells in the mammalian blastula to two layers—the inner cell mass and the trophoblast—results in the formation of the blastocyst.

Concept in Action

Visit the Virtual Human Embryo project at the Endowment for Human Development site to step through an interactive that shows the stages of embryo development, including micrographs and rotating 3-D images.

Gastrulation

The typical blastula is a ball of cells. The next stage in embryonic development is the formation of the body plan. The cells in the blastula rearrange themselves spatially to form three layers of cells. This process is called **gastrulation**. During gastrulation, the blastula folds upon itself to form the three layers of cells. Each of these layers is called a germ layer and each germ layer differentiates into different organ systems.

The three germs layers, shown in Figure 24.26, are the endoderm, the ectoderm, and the mesoderm. The ectoderm gives rise to the nervous system and the epidermis. The mesoderm gives rise to the muscle cells and connective tissue in the body. The endoderm gives rise to columnar cells found in the digestive system and many internal organs.

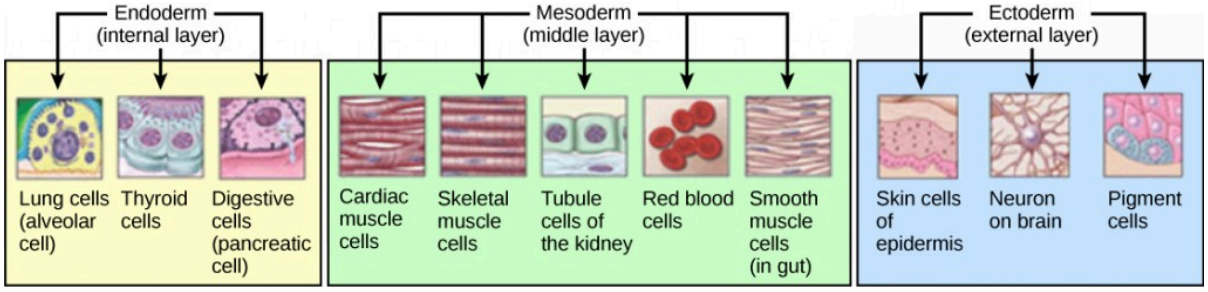


Figure 24.26. The three germ layers give rise to different cell types in the animal body. (credit: modification of work by NIH, NCBI)

Are Designer Babies in Our Future?

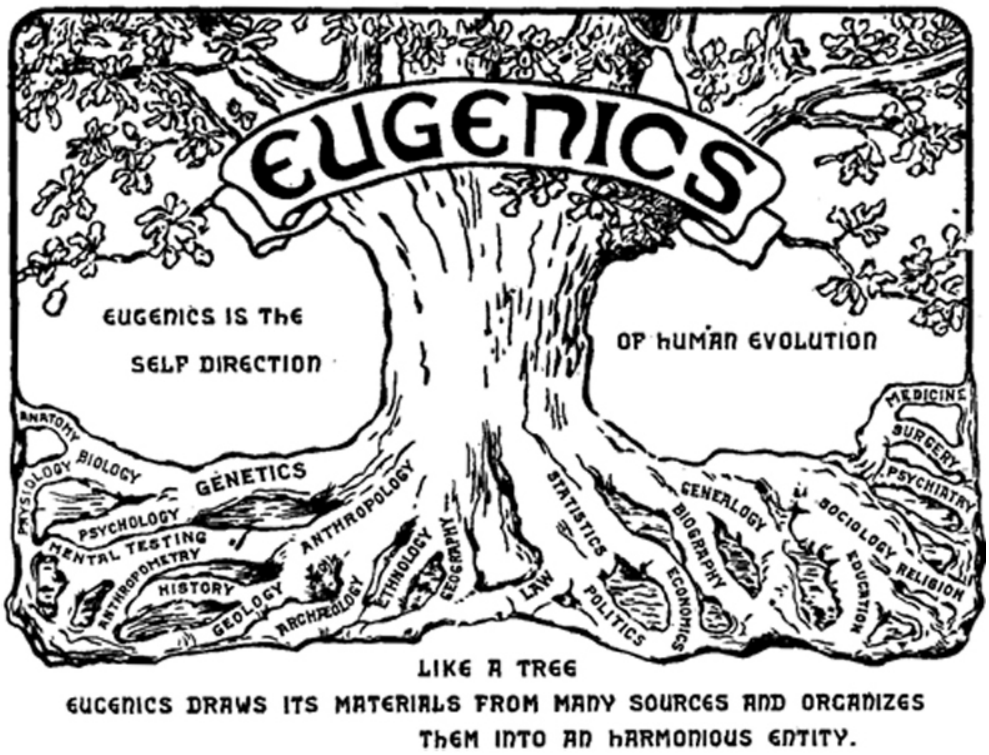


Figure 24.27. This logo from the Second International Eugenics Conference in New York City in September of 1921 shows how eugenics attempted to merge several fields of study with the goal of producing a genetically superior human race.

If you could prevent your child from getting a devastating genetic disease, would you do it? Would you select the sex of your child or select for their attractiveness, strength, or intelligence? How far would you go to maximize the possibility of resistance to disease? The genetic engineering of a human child, the production of "designer babies" with desirable phenotypic characteristics, was once a topic restricted to science fiction. This is the case no longer: science fiction is now overlapping into science fact. Many phenotypic choices for offspring are already available, with many more likely to be possible in the not too distant future. Which traits should be selected and how they should be selected are topics of much debate within the worldwide medical community. The ethical and moral line is not always clear or agreed upon, and some fear that modern reproductive technologies could lead to a new form of eugenics.

Eugenics is the use of information and technology from a variety of sources to improve the genetic makeup of the human race. The goal of creating genetically superior humans was quite prevalent (although controversial) in several countries during the early 20th century, but fell into disrepute when Nazi Germany developed an extensive eugenics program in the 1930's and 40's. As part of their program, the Nazis forcibly sterilized hundreds of thousands of the so-called "unfit" and killed tens of thousands of institutionally disabled people as part of a systematic program to develop a genetically superior race of Germans known as Aryans. Ever since, eugenic ideas have not been as publicly expressed, but there are still those who promote them.

Efforts have been made in the past to control traits in human children using donated sperm from men

with desired traits. In fact, eugenicist Robert Klark Graham established a sperm bank in 1980 that included samples exclusively from donors with high IQs. The “genius” sperm bank failed to capture the public’s imagination and the operation closed in 1999.

In more recent times, the procedure known as prenatal genetic diagnosis (PGD) has been developed. PGD involves the screening of human embryos as part of the process of *in vitro* fertilization, during which embryos are conceived and grown outside the mother’s body for some period of time before they are implanted. The term PGD usually refers to both the diagnosis, selection, and the implantation of the selected embryos.

In the least controversial use of PGD, embryos are tested for the presence of alleles which cause genetic diseases such as sickle cell disease, muscular dystrophy, and hemophilia, in which a single disease-causing allele or pair of alleles has been identified. By excluding embryos containing these alleles from implantation into the mother, the disease is prevented, and the unused embryos are either donated to science or discarded. There are relatively few in the worldwide medical community that question the ethics of this type of procedure, which allows individuals scared to have children because of the alleles they carry to do so successfully. The major limitation to this procedure is its expense. Not usually covered by medical insurance and thus out of reach financially for most couples, only a very small percentage of all live births use such complicated methodologies. Yet, even in cases like these where the ethical issues may seem to be clear-cut, not everyone agrees with the morality of these types of procedures. For example, to those who take the position that human life begins at conception, the discarding of unused embryos, a necessary result of PGD, is unacceptable under any circumstances.

A murkier ethical situation is found in the selection of a child’s sex, which is easily performed by PGD. Currently, countries such as Great Britain have banned the selection of a child’s sex for reasons other than preventing sex-linked diseases. Other countries allow the procedure for “family balancing”, based on the desire of some parents to have at least one child of each sex. Still others, including the United States, have taken a scattershot approach to regulating these practices, essentially leaving it to the individual practicing physician to decide which practices are acceptable and which are not.

Even murkier are rare instances of disabled parents, such as those with deafness or dwarfism, who select embryos via PGD to ensure that they share their disability. These parents usually cite many positive aspects of their disabilities and associated culture as reasons for their choice, which they see as their moral right. To others, to purposely cause a disability in a child violates the basic medical principle of *Primum non nocere*, “first, do no harm.” This procedure, although not illegal in most countries, demonstrates the complexity of ethical issues associated with choosing genetic traits in offspring.

Where could this process lead? Will this technology become more affordable and how should it be used? With the ability of technology to progress rapidly and unpredictably, a lack of definitive guidelines for the use of reproductive technologies before they arise might make it difficult for legislators to keep pace once they are in fact realized, assuming the process needs any government regulation at all. Other bioethicists argue that we should only deal with technologies that exist now, and not in some uncertain future. They argue that these types of procedures will always be expensive and rare, so the fears of eugenics and “master” races are unfounded and overstated. The debate continues.

Summary

The early stages of embryonic development begin with fertilization. The process of fertilization is tightly controlled to ensure that only one sperm fuses with one egg. After fertilization, the zygote undergoes cleavage to form the blastula. The blastula, which in some species is a hollow ball of cells, undergoes a process called gastrulation, in which the three germ layers form. The ectoderm gives rise to the nervous system and the epidermal skin cells, the mesoderm gives rise to the muscle cells and connective tissue in the body, and the endoderm gives rise to columnar cells and internal organs.

Exercises

1. Which of the following is false?
    1. The endoderm, mesoderm, ectoderm are germ layers.
    2. The trophoblast is a germ layer.
    3. The inner cell mass is a source of embryonic stem cells.
    4. The blastula is often a hollow ball of cells.
2. During cleavage, the mass of cells:
    1. increases
    2. decreases
    3. doubles with every cell division
    4. does not change significantly
3. What do you think would happen if multiple sperm fused with one egg?
4. Why do mammalian eggs have a small concentration of yolk, while bird and reptile eggs have a large concentration of yolk?

Answers

1. B
2. D
3. Multiple sperm can fuse with the egg, resulting in polyspermy. The resulting embryo is not genetically viable and dies within a few days.
4. Mammalian eggs do not need a lot of yolk because the developing fetus obtains nutrients from the mother. Other species, in which the fetus develops outside of the mother's body, such as occurs with birds, require a lot of yolk in the egg to nourish the embryo during development.

Glossary

acrosomal reaction
series of biochemical reactions that the sperm uses to break through the zona pellucida

blastocyst
structure formed when cells in the mammalian blastula separate into an inner and outer layer

gastrulation
process in which the blastula folds over itself to form the three germ layers

holoblastic
complete cleavage; takes place in cells with a small amount of yolk

inner cell mass
inner layer of cells in the blastocyst

meroblastic
partial cleavage; takes place in cells with a large amount of yolk

polyspermy
condition in which one egg is fertilized by multiple sperm

trophoblast
outer layer of cells in the blastocyst

zona pellucida
protective layer of glycoproteins on the mammalian egg

24.7. Organogenesis and Vertebrate Formation

Learning Objectives

By the end of this section, you will be able to:

- Describe the process of organogenesis
- Identify the anatomical axes formed in vertebrates

Gastrulation leads to the formation of the three germ layers that give rise, during further development, to the different organs in the animal body. This process is called **organogenesis**. Organogenesis is characterized by rapid and precise movements of the cells within the embryo.

Organogenesis

Organs form from the germ layers through the process of differentiation. During differentiation, the embryonic stem cells express specific sets of genes which will determine their ultimate cell type. For example, some cells in the ectoderm will express the genes specific to skin cells. As a result, these cells will differentiate into epidermal cells. The process of differentiation is regulated by cellular signaling cascades.

Scientists study organogenesis extensively in the lab in fruit flies (*Drosophila*) and the nematode *Caenorhabditis elegans*. *Drosophila* have segments along their bodies, and the patterning associated with the segment formation has allowed scientists to study which genes play important roles in organogenesis along the length of the embryo at different time points. The nematode *C.elegans* has roughly 1000 somatic cells and scientists have studied the fate of each of these cells during their development in the nematode life cycle. There is little variation in patterns of cell lineage between individuals, unlike in mammals where cell development from the embryo is dependent on cellular cues.

In vertebrates, one of the primary steps during organogenesis is the formation of the neural system. The ectoderm forms epithelial cells and tissues, and neuronal tissues. During the formation of the neural system, special signaling molecules called growth factors signal some cells at the edge of the ectoderm to become epidermis cells. The remaining cells in the center form the neural plate. If the signaling by growth factors were disrupted, then the entire ectoderm would differentiate into neural tissue.

The neural plate undergoes a series of cell movements where it rolls up and forms a tube called the **neural tube**, as illustrated in

Figure 24.28. In further development, the neural tube will give rise to the brain and the spinal cord.

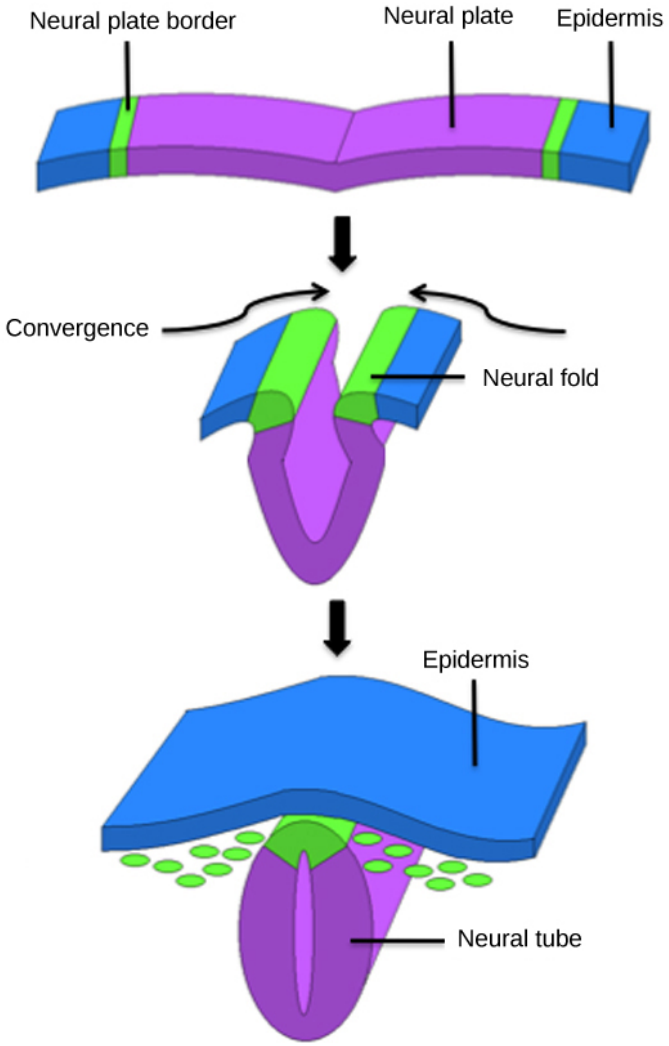


Figure 24.28. The central region of the ectoderm forms the neural tube, which gives rise to the brain and the spinal cord.

The mesoderm that lies on either side of the vertebrate neural tube will develop into the various connective tissues of the animal body. A spatial pattern of gene expression reorganizes the mesoderm into groups of cells called **somites** with spaces between them. The somites, illustrated in Figure 24.29 will further develop into the ribs, lungs, and segmental (spine) muscle. The mesoderm also forms a structure called the notochord, which is rod-shaped and forms the central axis of the animal body.

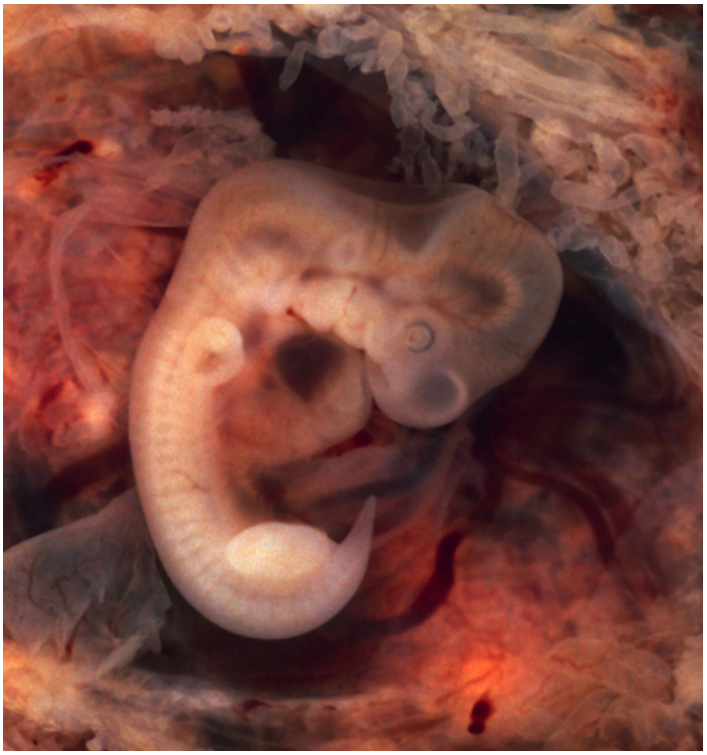


Figure 24.29. In this five-week old human embryo, somites are segments along the length of the body. (credit: modification of work by Ed Uthman)

Vertebrate Axis Formation

Even as the germ layers form, the ball of cells still retains its spherical shape. However, animal bodies have lateral-medial (left-right), dorsal-ventral (back-belly), and anterior-posterior (head-feet) axes, illustrated in Figure 24.30.

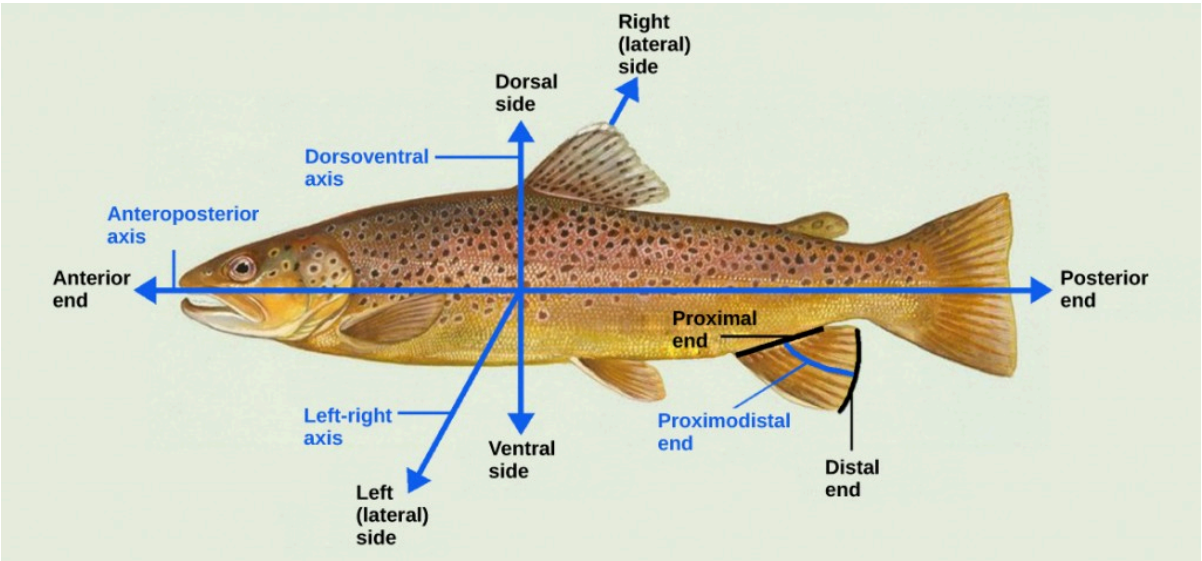


Figure 24.30. Animal bodies have three axes for symmetry. (credit: modification of work by NOAA)

How are these established? In one of the most seminal experiments ever to be carried out in developmental biology, Spemann and Mangold took dorsal cells from one embryo and transplanted them into the belly region of another embryo. They found that the transplanted embryo now had two notochords: one at the dorsal site from the original cells and another at the transplanted site. This suggested that the dorsal cells were genetically programmed to form the notochord and define the axis. Since then, researchers have identified many genes that are responsible for axis formation. Mutations in these genes leads to the loss of symmetry required for organism development.

Animal bodies have externally visible symmetry. However, the internal organs are not symmetric. For example, the heart is on the left side and the liver on the right. The formation of the central left-right axis is an important process during development. This internal asymmetry is established very early during development and involves many genes. Research is still ongoing to fully understand the developmental implications of these genes.

Summary

Organogenesis is the formation of organs from the germ layers. Each germ layer gives rise to specific tissue types. The first stage is the formation of the neural system in the ectoderm. The mesoderm gives rise to somites and the notochord. Formation of vertebrate axis is another important developmental stage.

Exercises

1. Which of the following gives rise to the skin cells?
    1. ectoderm
    2. endoderm
    3. mesoderm
    4. none of the above
2. The ribs form from the _____.
    1. notochord
    2. neural plate
    3. neural tube
    4. somites
3. Explain how the different germ layers give rise to different tissue types.
4. Explain the role of axis formation in development

Answers

1. A
2. D

3. Organs form from the germ layers through the process of differentiation. During differentiation, the embryonic stem cells express a specific set of genes that will determine their ultimate fate as a cell type. For example, some cells in the ectoderm will express the genes specific to skin cells. As a result, these cells will differentiate into epidermal cells. The process of differentiation is regulated by cellular signaling cascades.
4. Animal bodies have lateral-medial (left-right), dorsal-ventral (back-belly), and anterior-posterior (head-feet) axes. The dorsal cells are genetically programmed to form the notochord and define the axis. There are many genes responsible for axis formation. Mutations in these genes lead to the loss of symmetry required for organism development.

Glossary

morning sickness
condition in the mother during the first trimester; includes feelings of nausea

neural tube
tube-like structure that forms from the ectoderm and gives rise to the brain and spinal cord

oogenesis
process of producing haploid eggs

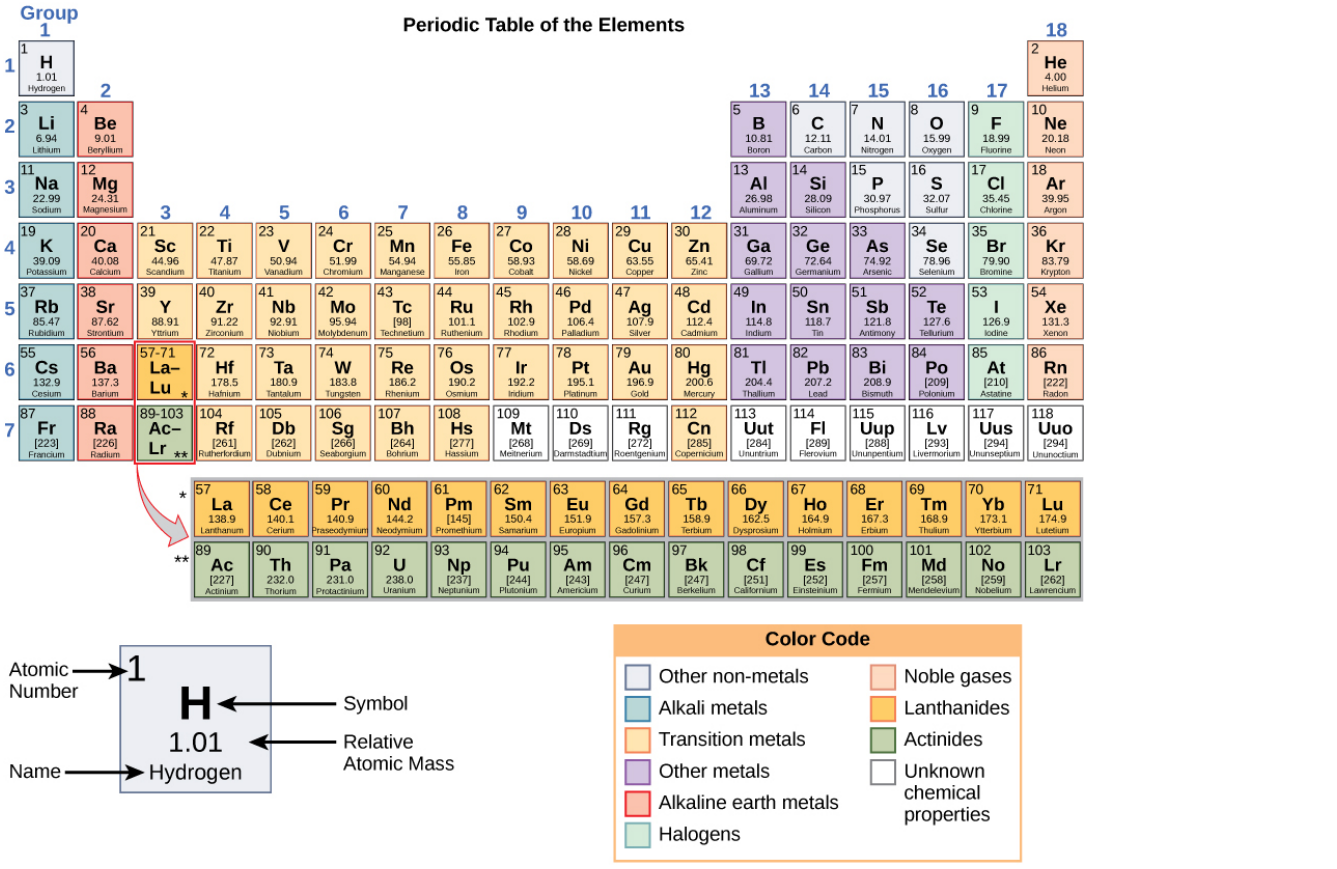
organogenesis
process of organ formation

oviduct
(also, fallopian tube) muscular tube connecting the uterus with the ovary area

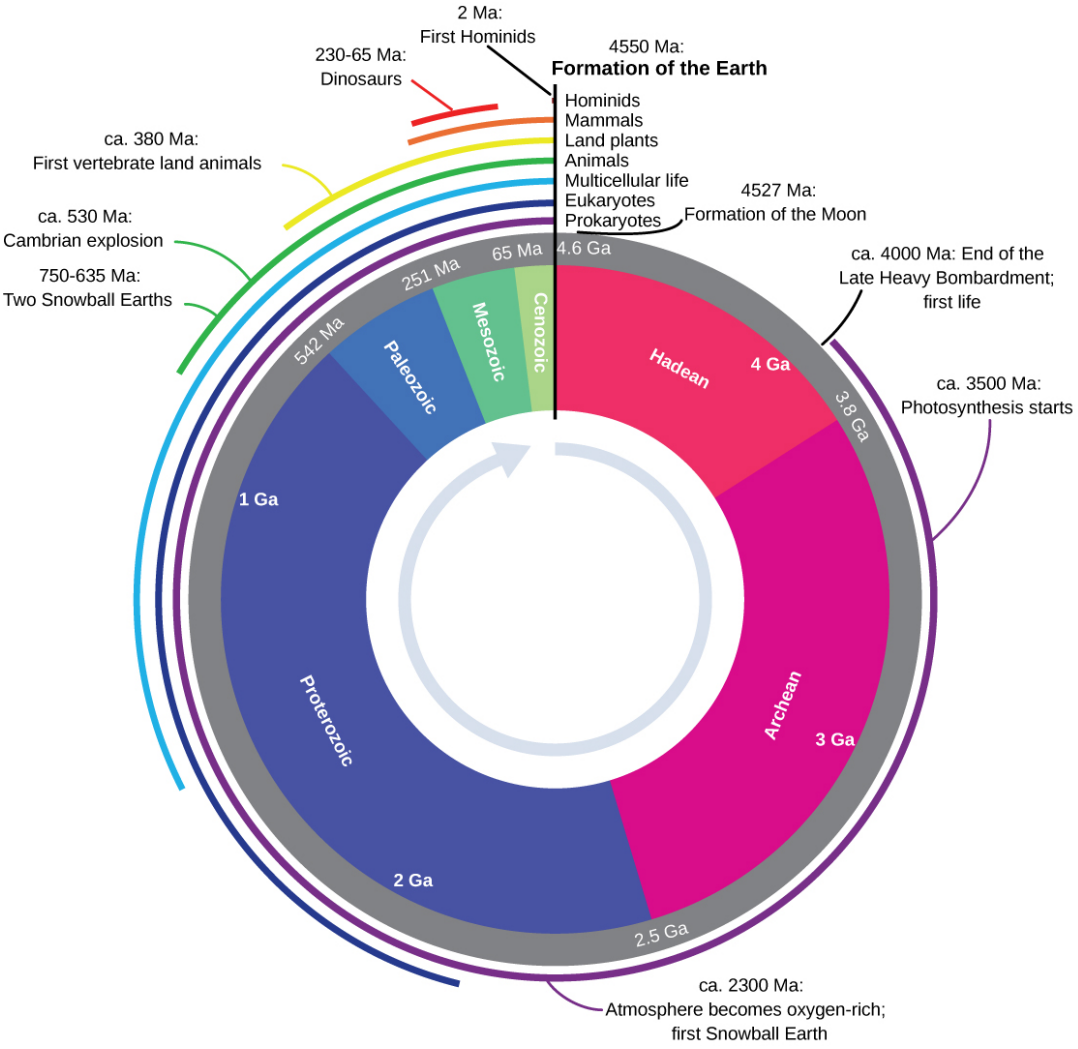
somite
group of cells separated by small spaces that form from the mesoderm and give rise to connective tissue

Appendix

Periodic Table of the Elements

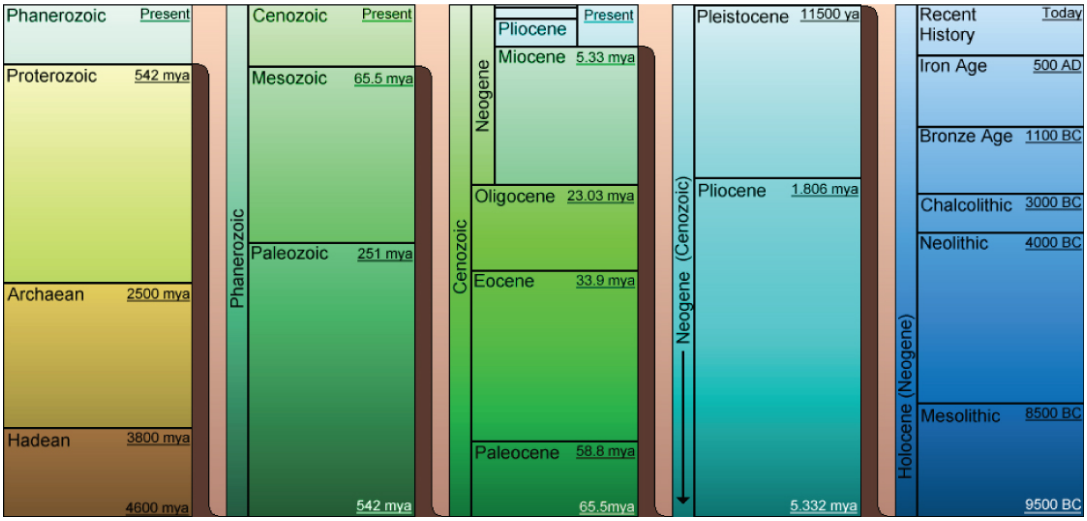


Geological Time Clock



Geological Time Chart

(credit: Richard S. Murphy, Jr.)



Measurements and the Metric System

Measurements and the Metric System

Measurement	Unit	Abbreviation	Metric Equivalent	Approximate Standard Equivalent
Length	nanometer	nm	1 nm = 10^{-9} m	• 1 mm = 0.039 inch • 1 cm = 0.394 inch • 1 m = 39.37 inches • 1 m = 3.28 feet • 1 m = 1.093 yards • 1 km = 0.621 miles
	micrometer	µm	1 µm = 10^{-6} m	
	millimeter	mm	1 mm = 0.001 m	
	centimeter	cm	1 cm = 0.01 m	
	meter	m	• 1 m = 100 cm • 1 m = 1000 mm	
	kilometer	km	1 km = 1000 m	
Mass	microgram	µg	1 µg = 10^{-6} g	• 1 g = 0.035 ounce • 1 kg = 2.205 pounds
	milligram	mg	1 mg = 10^{-3} g	
	gram	g	1 g = 1000 mg	
	kilogram	kg	1 kg = 1000 g	
Volume	microliter	µl	1 µl = 10^{-6} l	• 1 ml = 0.034 fluid ounce • 1 l = 1.057 quarts • 1 kl = 264.172 gallons
	milliliter	ml	1 ml = 10^{-3} l	
	liter	l	1 l = 1000 ml	
	kiloliter	kl	1 kl = 1000 l	
Area	square centimeter	cm^2	1 cm^2 = 100 mm^2	• 1 cm^2 = 0.155 square inch • 1 m^2 = 10.764 square feet • 1 m^2 = 1.196 square yards • 1 ha = 2.471 acres
	square meter	m^2	1 m^2 = 10,000 cm^2	
	hectare	ha	1 ha = 10,000 m^2	
Temperature	Celsius	°C	—	1 °C = 5/9 × (°F − 32)

PowerPoints

- Chapter 1 PowerPoint
- Chapter 2 PowerPoint
- Chapter 3 PowerPoint
- Chapter 4 PowerPoint
- Chapter 5 PowerPoint
- Chapter 6 PowerPoint
- Chapter 7 PowerPoint
- Chapter 8 PowerPoint
- Chapter 9 PowerPoint
- Chapter 10 PowerPoint
- Chapter 11 PowerPoint
- Chapter 12 PowerPoint
- Chapter 13 PowerPoint
- Chapter 14 PowerPoint
- Chapter 15 PowerPoint
- Chapter 16 PowerPoint
- Chapter 17 PowerPoint
- Chapter 18 PowerPoint
- Chapter 19 PowerPoint
- Chapter 20 PowerPoint
- Chapter 21 PowerPoint
- Chapter 22 PowerPoint
- Chapter 23 PowerPoint
- Chapter 24 PowerPoint

About the Authors

Charles Molnar

I have been fortunate to have been at Camosun College for 25 years, which has allowed me to live in a beautiful place, grow in many ways and be nourished by that which brings me happiness—teaching. While the University of Alberta is where my degrees came from, Camosun College is my alma mater. I am pleased to offer this work to all, and especially to my father, Dr. George Molnar, who was and is my scientific inspiration and my life's model.

Jane Gair

I have always loved learning and admired the many teachers in my life. They have been my role models and I am proud to call myself a teacher now. I am blessed to be able to teach for three institutions in British Columbia: Camosun College, the University of Victoria, and the University of British Columbia—my alma mater. It gives me great pleasure to know that the many students that I encounter in the classroom and beyond will use this textbook.

Versioning History

This page provides a record of edits and changes made to this book since its initial publication in the B.C. Open Textbook Collection. Whenever edits or updates are made, we make the required changes in the text and provide a record and description of those changes here. If the change is minor, the version number increases by 0.01. However, if the edits involve substantial updates, the version number goes up to the next full number. The files on our website always reflect the most recent version, including the print-on-demand copy.

If you find an error in this book, please fill out the Report an Open Textbook Error form.

Version	Date	Change	Details
1.01	May 1, 2015	Book added to the B.C. Open Textbook collection.	
1.02	November 21, 2017	• Fixed blue text • Reformatted exercise questions	Some <a id=””> tags in the book were causing large sections of text to turn blue. These tags were removed. In addition, all of the exercise questions were edited to ensure a consistent format. This involved putting the content into ordered lists and grouping questions and answers together. Note: We are aware that there is some text missing in Chapter 20.4, that Chapter 14.1 is missing exercise questions, and that Chapter 14.2 and 14.3 have the same exercise questions. The author has been informed of these issues and we will update the book once we receive the corrections.
1.03	September 17, 2018	Image 2.15 replaced.	Original Figure 2.15 image was a duplicate of Figure 2.17. Replaced Figure 2.15 with the correct image.
1.04	June 13, 2019	Updated the book’s theme.	The styles of this book have been updated, which may affect the page numbers of the PDF and print copy.
2.01	March 3, 2021	Added H5P activities and improved book structure and formatting.	80 H5P activities were embedded throughout the book, book structure was revised to improve navigation (but the order of all chapters remains the same), all PowerPoint files were compiled in the back matter of the book, and front matter and metadata was updated.
2.02	January 5, 2022	Updated image attributions.	Corrected figure numbering and attribution statements in Chapter 2.3 Biological Molecules.
2.03	May 31, 2022	Added missing text.	Added a missing key term in the text under the header “Transport of Carbon Dioxide in the Blood” in Chapter 20.4 Transport of Gases in Human Bodily Fluids.